ESSENTIAL REFERENCE

Industrial Engineering
Concepts, Methodologies, Tools, and Applications

Volume I

Information Resources Management Association

Industrial Engineering:

Concepts, Methodologies, Tools and Applications

Information Resources Management Association
USA

Volume I

ENGINEERING SCIENCE REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Book Production Manager: Jennifer Romanchak
Publishing Systems Analyst: Adrienne Freeland
Assistant Acquisitions Editor: Kayla Wolfe
Development Manager: Joel Gamon
Development Editor: Chris Wozniak
Assistant Production Editor: Deanna Jo Zombro
Cover Design: Nick Newcomer

Published in the United States of America by
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Industrial engineering : concepts, methodologies, tools, and applications / Information Resources Management Association, editor.
v. cm.
Includes bibliographical references and index.
ISBN 978-1-4666-1945-6 (hardcover) -- ISBN 978-1-4666-1946-3 (ebook) -- ISBN 978-1-4666-1947-0 (print & perpetual access) 1. Industrial engineering. 2. Industrial engineering--Case studies. I. Information Resources Management Association.
T56.I43 2013
620--dc23

2012023210

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

List of Contributors

Table of Contents

Volume I

Section 1
Fundamental Concepts and Theories

This section serves as a foundation for this exhaustive reference tool by addressing underlying principles essential to the understanding of Industrial Engineering. Chapters found within these pages provide an excellent framework in which to position Industrial Engineering within the field of information science and technology. Insight regarding the critical incorporation of global measures into Industrial Engineering is addressed, while crucial stumbling blocks of this field are explored. With 10 chapters comprising this foundational section, the reader can learn and chose from a compendium of expert research on the elemental theories underscoring the Industrial Engineering discipline.

Section 2
Development and Design Methodologies

This section provides in-depth coverage of conceptual architecture frameworks to provide the reader with a comprehensive understanding of the emerging developments within the field of Industrial Engineering. Research fundamentals imperative to the understanding of developmental processes within Industrial Engineering are offered. From broad examinations to specific discussions on methodology, the research found within this section spans the discipline while offering detailed, specific discussions. From basic designs to abstract development, these chapters serve to expand the reaches of development and design technologies within the Industrial Engineering community. This section includes 14 contributions from researchers throughout the world on the topic of Industrial Engineering.

Section 3
Tools and Technologies

This section presents an extensive coverage of various tools and technologies available in the field of Industrial Engineering that practitioners and academicians alike can utilize to develop different techniques. These chapters enlighten readers about fundamental research on the many tools facilitating the burgeoning field of Industrial Engineering. It is through these rigorously researched chapters that the reader is provided with countless examples of the up-and-coming tools and technologies emerging from the field of Industrial Engineering. With 14 chapters, this section offers a broad treatment of some of the many tools and technologies within the Industrial Engineering field.

Volume II

Section 4
Utilization and Application

This section discusses a variety of applications and opportunities available that can be considered by practitioners in developing viable and effective Industrial Engineering programs and processes. This section includes 14 chapters that review topics from case studies in Cyprus to best practices in Africa and ongoing research in the United States. Further chapters discuss Industrial Engineering in a variety of settings (air travel, education, gaming, etc.). Contributions included in this section provide excellent coverage of today's IT community and how research into Industrial Engineering is impacting the social fabric of our present-day global village.

Section 6
Managerial Impact

This section presents contemporary coverage of the social implications of Industrial Engineering, more specifically related to the corporate and managerial utilization of information sharing technologies and applications, and how these technologies can be extrapolated to be used in Industrial Engineering. Core ideas such as service delivery, gender evaluation, public participation, and other determinants that affect the intention to adopt technological innovations in Industrial Engineering are discussed. Equally as crucial, chapters within this section discuss how leaders can utilize Industrial Engineering applications to get the best outcomes from their shareholders and their customers.

Volume III

Section 7
Critical Issues

This section contains 13 chapters, giving a wide variety of perspectives on Industrial Engineering and its implications. Such perspectives include reading in privacy, gender, ethics, and several more. The section also discusses new ethical considerations within social constructivism and gender gaps. Within the chapters, the reader is presented with an in-depth analysis of the most current and relevant issues within this growing field of study. Crucial questions are addressed and alternatives offered, and topics discussed such as creative regions in Europe, ethos as an enabler of organizational knowledge creation, and design of manufacturing cells based on graph theory.

Preface

The constantly changing landscape of Industrial Engineering makes it challenging for experts and practitioners to stay informed of the field's most up-to-date research. That is why Information Science Reference is pleased to offer this three-volume reference collection that will empower students, researchers, and academicians with a strong understanding of critical issues within Industrial Engineering by providing both broad and detailed perspectives on cutting-edge theories and developments. This reference is designed to act as a single reference source on conceptual, methodological, technical, and managerial issues, as well as provide insight into emerging trends and future opportunities within the discipline.

Industrial Engineering: Concepts, Methodologies, Tools and Applications is organized into eight distinct sections that provide comprehensive coverage of important topics. The sections are: (1) Fundamental Concepts and Theories, (2) Development and Design Methodologies, (3) Tools and Technologies, (4) Utilization and Application, (5) Organizational and Social Implications, (6) Managerial Impact, (7) Critical Issues, and (8) Emerging Trends. The following paragraphs provide a summary of what to expect from this invaluable reference tool.

Section 1, **Fundamental Concepts and Theories**, serves as a foundation for this extensive reference tool by addressing crucial theories essential to the understanding of Industrial Engineering. Introducing the book is "*Defining, Teaching, and Assessing Engineering Design Skills*" by Nikos J. Mourtos, a great foundation laying the groundwork for the basic concepts and theories that will be discussed throughout the rest of the book. Another chapter of note in Section 1 is titled "*Integrating 'Designerly' Ways with Engineering Science*" by Ian de Vere and Gavin Melles, which discusses the novel techniques of adding aspects of design science into the stricter roles of engineering practices. Section 1 concludes, and leads into the following portion of the book with a nice segue chapter, "*Tracing the Implementation of Non-Functional Requirements*," by Stephan Bode and Matthias Riebisch. Where Section 1 leaves off with fundamental concepts, Section 2 discusses architectures and frameworks in place for Industrial Engineering.

Section 2, **Development and Design Methodologies**, presents in-depth coverage of the conceptual design and architecture of Industrial Engineering, focusing on aspects including parametric design, service design, fuzzy logic, control modeling, supply chain systems, and many more topics. Opening the section is "*Learning Parametric Designing*" by Marc Aurel Schnabel. This section is vital for developers and practitioners who want to measure and track the progress of Industrial Engineering on a through the multiple lens of parametric design. Through case studies, this section lays excellent groundwork for later sections that will get into present and future applications for Industrial Engineering, including, of note: "*Decision Support Framework for the Selection of a Layout Type*" by Jannes Slomp and Jos A.C. Bokhorst, and "*Internal Supply Chain Integration*" by Virpi Turkulainen. The section concludes with an excellent work by Mousumi Debnath and Mukeshwar Pandey, titled "*Enhancing Engineering Education Learning Outcomes Using Project-Based Learning*."

Section 3, **Tools and Technologies**, presents extensive coverage of the various tools and technologies used in the implementation of Industrial Engineering. Section 3 begins where Section 2 left off, though this section describes more concrete tools at place in the modeling, planning, and applications of Industrial Engineering. The first chapter, "*Semantic Technologies in Motion*," by Ricardo Colomo-Palacios, lays a framework for the types of works that can be found in this section, a perfect resource for practitioners looking for the fundamentals of the types of semantic technologies currently in practice in Industrial Engineering. Section 3 is full of excellent chapters like this one, including such titles as "*Optimization and Mathematical Programming to Design and Planning Issues in Cellular Manufacturing Systems under Uncertain Situations*," "*Multi-Modal Assembly-Support System for Cellular Manufacturing*," and "*An Estimation of Distribution Algorithm for Part Cell Formation Problem*" to name a few. Where Section 3 described specific tools and technologies at the disposal of practitioners, Section 4 describes successes, failures, best practices, and different applications of the tools and frameworks discussed in previous sections.

Section 4, **Utilization and Application**, describes how the broad range of Industrial Engineering efforts has been utilized and offers insight on and important lessons for their applications and impact. Section 4 includes the widest range of topics because it describes case studies, research, methodologies, frameworks, architectures, theory, analysis, and guides for implementation. Topics range from serios games, enterprise resource planning, and crisis management, to air travel development and design. The first chapter in the section is titled "*Using Serious Games for Collecting and Modeling Human Procurement Decisions in a Supply Chain Context*," which was written by Souleiman Naciri, Min-Jung Yoo, and Rémy Glardon. The breadth of topics covered in the chapter is also reflected in the diversity of its authors, from countries all over the globe, including Germany, Slovenia, Norway, Hong Kong, Malaysia, Brazil, Cyprus, Turkey, the United States, and more. Section 4 concludes with an excellent view of a case study in a new program, "*UB1-HIT Dual Master's Programme,*" by David Chen, Bruno Vallespir, Jean-Paul Bourrieres, and Thecle Alix.

Section 5, **Organizational and Social Implications**, includes chapters discussing the organizational and social impact of Industrial Engineering. The section opens with "*Process Innovation with Ambient Intelligence (AmI) Technologies in Manufacturing SMEs*" by Kathryn J. Hayes and Ross Chapman. Where Section 4 focused on the broad, many applications of Industrial Engineering technology, Section 5 focuses exclusively on how these technologies affect human lives, either through the way they interact with each other, or through how they affect behavioral/workplace situations. Other interesting chapters of note in Section 5 include "*Group Decision Making for Advanced Manufacturing Technology Selection Using the Choquet Integral*" by Cengiz Kahraman, Selçuk Çebi, and Ihsan Kaya, and "*Direct Building Manufacturing of Homes with Digital Fabrication*" by Lawrence Sass. Section 5 concludes with a fascinating study of a new development in Industrial Engineering, in "*Firm-Specific Factors and the Degree of Innovation Openness*" by Valentina Lazzarotti, Raffaella Manzini, and Luisa Pellegrini.

Section 6, **Managerial Impact**, presents focused coverage of Industrial Engineering as it relates to effective uses of offshoring, network marketing, knowledge management, e-government, knowledge dissemination, and many more utilities. This section serves as a vital resource for developers who want to utilize the latest research to bolster the capabilities and functionalities of their processes. The section begins with "*Offshoring Process*," a great look into whether or not offshoring practices could help a given business, alongside best practices and some new trends in the field. The 13 chapters in this section offer unmistakable value to managers looking to implement new strategies that work at larger bureaucratic levels. The section concludes with "*Research Profiles*" by Gretchen Jordan, Jonathon Mote, and Jerald Hage. Where Section 6 leaves off, section seven picks up with a focus on some of the more content-theoretical material of this compendium.

Section 7, **Critical Issues**, presents coverage of academic and research perspectives on Industrial Engineering tools and applications. The section begins with "*Cultural Models and Variations*" by Yongjiang Shi and Zheng Liu. Other issues covered in detail in Section 7 include design paradigns, knowledge dynamics, layout structuring, design ethos, and much more. The section concludes with "*Engineer-to-Order*" by Ephrem Eyob and Richard Addo-Tenkorang, a great transitional chapter between Sections 7 and 8 because it examines an important trend going into the future of the field. The last chapter manages to show a theoretical look into future and potential technologies, a topic covered in more detail in Section 8.

Section 8, **Emerging Trends**, highlights areas for future research within the field of Industrial Engineering, opening with "*Advanced Technologies for Transient Faults Detection and Compensation*" by Matteo Sonza Reorda, Luca Sterpone, and Massimo Violante. Section 8 contains chapters that look at what might happen in the coming years that can extend the already staggering amount of applications for Industrial Engineering. Other chapters of note include "*Embedded RFID Solutions Challenges for Product Design and Development*" and "*Green Computing as an Ecological Aid in Industry*." The final chapter of the book looks at an emerging field within Industrial Engineering, in the excellent contribution, "*Zero-Downtime Reconfiguration of Distributed Control Logic in Industrial Automation and Control*" by Thomas Strasser and Alois Zoitl.

Although the primary organization of the contents in this multi-volume work is based on its eight sections, offering a progression of coverage of the important concepts, methodologies, technologies, applications, social issues, and emerging trends, the reader can also identify specific contents by utilizing the extensive indexing system listed at the end of each volume. Furthermore to ensure that the scholar, researcher, and educator have access to the entire contents of this multi volume set as well as additional coverage that could not be included in the print version of this publication, the publisher will provide unlimited multi-user electronic access to the online aggregated database of this collection for the life of the edition, free of charge when a library purchases a print copy. This aggregated database provides far more contents than what can be included in the print version, in addition to continual updates. This unlimited access, coupled with the continuous updates to the database ensures that the most current research is accessible to knowledge seekers.

As a comprehensive collection of research on the latest findings related to using technology to providing various services, *Industrial Engineering: Concepts, Methodologies, Tools and Applications*, provides researchers, administrators and all audiences with a complete understanding of the development of applications and concepts in Industrial Engineering. Given the vast number of issues concerning usage, failure, success, policies, strategies, and applications of Industrial Engineering in countries around the world, *Industrial Engineering: Concepts, Methodologies, Tools and Applications* addresses the demand for a resource that encompasses the most pertinent research in technologies being employed to globally bolster the knowledge and applications of Industrial Engineering.

Section 1
Fundamental Concepts and Theories

This section serves as a foundation for this exhaustive reference tool by addressing underlying principles essential to the understanding of Industrial Engineering. Chapters found within these pages provide an excellent framework in which to position Industrial Engineering within the field of information science and technology. Insight regarding the critical incorporation of global measures into Industrial Engineering is addressed, while crucial stumbling blocks of this field are explored. With 10 chapters comprising this foundational section, the reader can learn and chose from a compendium of expert research on the elemental theories underscoring the Industrial Engineering discipline.

Chapter 1
Defining, Teaching, and Assessing Engineering Design Skills

Nikos J. Mourtos
San Jose State University, USA

ABSTRACT

The paper discusses a systematic approach for defining, teaching, and assessing engineering design skills. Although the examples presented in the paper are from the field of aerospace engineering, the principles apply to engineering design in general. What makes the teaching of engineering design particularly challenging is that the necessary skills and attributes are both technical and non-technical and come from the cognitive as well as the affective domains. Each set of skills requires a different approach to teach and assess. Implementing a variety of approaches for a number of years at SJSU has shown that it is just as necessary to teach affective skills, as it is to teach cognitive skills. As one might expect, each set of skills presents its own challenges.

INTRODUCTION

Design is the heart of engineering practice. In fact, many engineering experts consider design as being synonymous with engineering. Yet engineering schools have come under increasing criticism after World War II because they have overemphasized analytical approaches and engineering science at the expense of hands-on, design skills (Seely, 1999; Petrosky, 2000). As the editor of Machine Design put it, *schools are being charged with not responding to industry needs for hands-on design talent, but instead are grinding out legions of research scientists* (Curry, 1991).

In response to this criticism and to increase student retention, many engineering schools, including SJSU, introduce design at the freshman level to excite students about engineering. Freshman design also helps students put into perspective the entire curriculum, by viewing each subject as

DOI: 10.4018/978-1-4666-1945-6.ch001

a necessary tool in the design process. Design is also globally dispersed in a variety of junior and senior level courses in the form of mini design projects and is finally experienced in a more realistic setting in a two-semester, senior design capstone experience.

The paper first attempts to provide a comprehensive definition of design skills. Subsequently, it presents a model for curriculum design that addresses these skills. Lastly, it presents ideas for assessing student competence in design. What makes teaching engineering design particularly challenging is that the necessary skills and attributes are technical as well as non-technical, and come from the cognitive as well as the affective domains. For example, the ability to define "real world" problems in practical (engineering) terms, to investigate and evaluate prior solutions, and to develop constraints and criteria for evaluation are technical skills, while the ability to communicate the results of a design, to work in teams, and decide on the best course of action when a decision has ethical implications are non-technical skills. Most technical skills are cognitive, however, there are several skills from the affective domain as well, such as the willingness to spend time reading, gathering information and defining the problem, and the willingness to risk and cope with ambiguity, to welcome change and manage stress. All these skills, technical and non-technical, cognitive and affective are essential for engineers, yet each requires a different approach to teach and assess.

DEFINING ENGINEERING DESIGN SKILLS

What is Engineering?

To define the skills necessary for design engineers we need to start with the definition of engineering itself. Nicolai (1988) defines engineering as *the design of a commodity for the benefit of mankind.* Obviously, the word *design* is key to the definition of engineering. Engineers design things in their attempt to solve everyday problems and improve the quality of our lives. As Theodore Von Karman put it: *A scientist discovers that which exists. An engineer creates that which never was*.

What is Design?

The next step in our search for design skills is to define design itself.

"Design is a process through which one creates and transforms ideas and concepts into a product that satisfies certain requirements and constraints."

Design requirements are usually technical and describe the performance expectations of the product, as specified by the customer or a perceived need. For example, a new passenger airplane may have mission requirements such as:

- A range of 3,000 km (i.e., the distance it will be able to fly without refueling).
- A payload of 100 passengers (i.e., the number of passengers along with their luggage it will be able to carry).
- A flight speed of 750 km/hr at a cruise altitude of 10 km.
- A takeoff field length of 1,500 m at standard sea level conditions.

The performance requirements specified by an airline (the customer), however, are not the only technical requirements that a passenger airplane must meet. To be certified, the plane must also satisfy additional airworthiness requirements. For example, FAR 25.121 part(b), refers to the ability of the plane to climb with one engine inoperative and requires that:

- In the takeoff configuration with the landing gear fully retracted but without ground effect the airplane must be able to maintain

> a steady climb gradient of at least 2.4% for two-engine airplanes, 2.7% for three-engine airplanes, and 3% for four-engine airplanes at a climb speed that is also specified and known as V2 (Flightsim Aviation Zone, 2010).

Such airworthiness requirements often prove to be more challenging than the original performance requirements specified by the customer. Additional design requirements, not specified by the customer, are not unique to aerospace engineering. For example, civil and architectural engineers must satisfy building code requirements, usually set by cities or countries.

The definition of design also mentions constraints. Constraints are sometimes difficult to distinguish from requirements. They may be viewed as limitations stated in regards to materials, cost, environmental factors, etc. For example, the Hughes H-4 Hercules aircraft, the largest flying boat ever built, was made out of wood because of wartime restrictions on the use of aluminum (Wikipedia, 2011). Another example is the noise standards for transport aircraft (Flightsim Aviation Zone, 2010).

In summary, design engineers must satisfy technical requirements, as specified by the customer and possibly additional technical requirements related to safety. Furthermore they must be concerned with the broader impact of their designs to individuals, the society, and the environment. This has become increasingly more important in our interconnected, globalized world.

Pink (2005) adds yet another challenge to engineering design, one that relates to aesthetics. He argues that because of the 'abundance' of products we have come to expect in the 21st century, the lower manufacturing cost in many countries, and the fact that many engineering tasks can now be automated, *it is no longer enough to create a product that's reasonably priced and adequately functional. It must also be beautiful, unique, and meaningful.* This requirement adds a new dimension to engineering design, a dimension that has much in common with the creative arts.

The Engineering Design Process

The next step in our search for design skills is to look at the engineering design process. Figure 1 is an attempt to illustrate this iterative process, as it takes place in our brain (Nicolai, 1998).

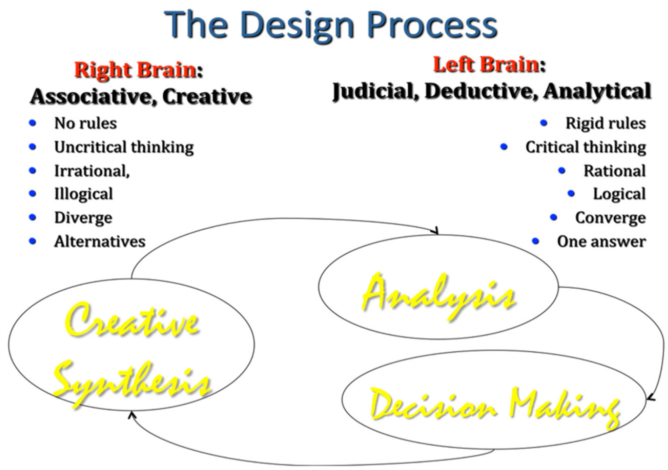
Design begins with brainstorming of ideas. This takes place in the right (creative) part of the brain. There are virtually no rules in generating these ideas. In fact, it is desirable to come up with as many ideas as possible and allow for "wild" ideas as well as conventional ones. While brainstorming, the right brain tends to be holistic, intuitive, and highly nonlinear (i.e., it jumps around). It sees things in their context as well as metaphorically, recognizes patterns, focuses on relationships between the various parts and cares about aesthetics.

Subsequently, each idea is evaluated in the left (analytical) part of the brain under very rigid rules. The left brain acts as a filter on the ideas generated, deciding which ones are viable under the current rules and which ones are not. The left brain tends to be logical, sequential, computer-like. It sees things literally and focuses on categories.

As Figure 1 illustrates, the design process involves an iterative cycling through a sequence that involves creative, imaginative exploration, objective analytical evaluation, and finally making a decision. It is this context, known also as convergent–divergent thinking (Nicolai, 1998), in which one should look for the skills and attributes necessary for a good design engineer.

But there is more to the iterative nature of engineering design than the interchange between the right and the left brain illustrated in Figure 1; iteration is also necessary because of the open-ended nature of design. It is simply not possible to follow a linear, step-by-step process to arrive at a single answer or a unique product that meets our need. First of all, design requires numerous

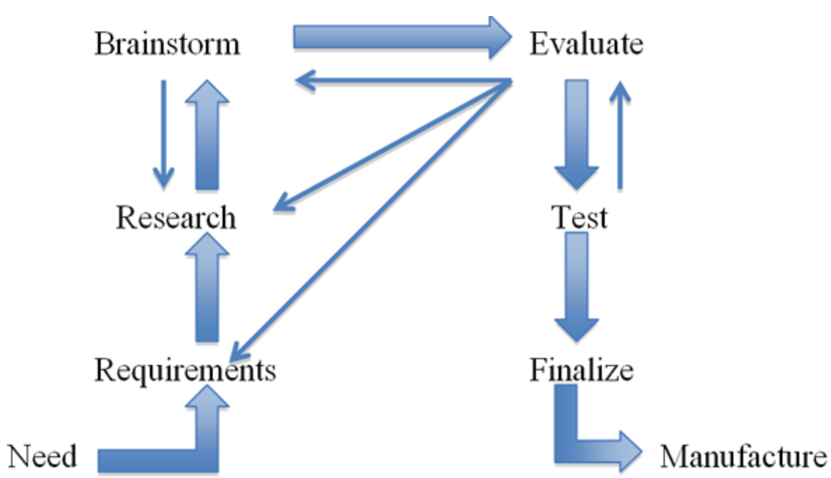
Figure 1. The engineering design process: an iteration between creative synthesis and analytical evaluation (adapted from Nicolai, 1998)



assumptions because there are always so many unknowns. Some of these assumptions may be proven wrong down the road, requiring us to go back, make changes, and repeat our calculations, hence the need for iteration. The non-unique nature of design becomes obvious when one looks at the multitude of products available in the market to address a given need.

Figure 2 illustrates the engineering design process. Engineering design begins with identifying a need. This need is articulated in terms of specific technical requirements that the product must meet. Following this design specification engineers research existing solutions to the problem before proposing any new ones. Brainstorming is the most creative part in the design process. The members of the design team who brainstorm typically bring various perspectives and expertise to the problem. The goal is to create as many ideas as possible, including unusual and wild ones. To achieve this goal, participants are not allowed to criticize any ideas put forth. Rather, to create synergy, they

Figure 2. The engineering design process: From identifying a need to production



are encouraged to build on others' ideas. After brainstorming the group selects two or three of these ideas to move forward with evaluation. Each proposed concept is analyzed systematically using appropriate engineering science in an effort to prove its feasibility and functionality. Hopefully, at least one of these concepts will prove feasible through analysis. A model is then built for actual testing. The tests will hopefully validate one of the proposed concepts, at which point the design is finalized and goes into production.

Design also requires compromise because requirements often conflict with each other. For example, to provide comfort for airplane passengers one needs a large cross-sectional area. But a large cross-sectional area results in greater drag and compromised fuel efficiency, especially at high speeds. A successful aircraft designer must decide where to draw the line between these two conflicting requirements.

Skills and Attributes of Design Engineers

Clearly, engineering design is a very complex process and as such, it requires several, very different from each other, sets of skills. These are briefly discussed in the following sub-sections.

Analytical Skills

The right-hand side of Figure 1 attests to the need for traditional engineering analytical skills: solid fundamentals in mathematics, physical science (e.g., physics, chemistry, etc.), and engineering science (e.g., fluid mechanics, thermodynamics, dynamics, etc.). The need for such skills has been articulated in the desired attributes of a global engineer (The Boeing Company & Rensselaer Polytechnic Institute, 1997), as well as in ABET EC 2000, Outcome 3a (Engineering Accreditation Commission):

"A good grasp of engineering science fundamentals, including: mechanics and dynamics, mathematics (including statistics), physical and life sciences, and information science/technology

An ability to apply knowledge of mathematics, science, and engineering"

Open-Ended Problem Solving Skills

Design skills build upon open-ended problem solving skills. Outcome 3e of ABET EC 2000 (Engineering Accreditation Commission) highlights the need for such skills when it states that engineering graduates must be able to *identify* and *formulate* engineering problems in addition to being able to solve such problems.

Students who are open-ended problem solvers exhibit the attributes listed below (Woods, 1997). Mourtos, Okamoto, and Rhee (2004) classified these attributes according to the various levels of Bloom's taxonomy of educational objectives in the cognitive and the affective domains (Bloom, 1984; Bloom, Karthwohl, & Massia, 1984):

a. Are willing to spend time reading, gathering information and defining the problem (Affective)
b. Use a process, as well as a variety of tactics and heuristics to tackle problems (Cognitive)
c. Monitor their problem-solving process and reflect upon its effectiveness (Affective and Cognitive)
d. Emphasize accuracy rather than speed (Affective and Cognitive)
e. Write down ideas and create charts / figures, while solving a problem (Affective and Cognitive)
f. Are organized and systematic (Affective)
g. Are flexible (keep options open, can view a situation from different perspectives / points of view) (Affective)
h. Draw on the pertinent subject knowledge and objectively and critically assess the quality,

accuracy, and pertinence of that knowledge / data (Cognitive)

i. Are willing to risk and cope with ambiguity, welcoming change and managing stress (Affective)

j. Use an overall approach that emphasizes fundamentals rather than trying to combine various memorized sample solutions (Cognitive)

It is interesting to note that the need for flexibility (attribute g) is also established as a desired attribute for a global engineer in a context much broader than engineering problem solving (The Boeing Company & Rensselaer Polytechnic Institute, 1997):

"Flexibility: the ability and willingness to adapt to rapid and/or major change."

The observation that some of these attributes are associated with the affective domain suggests that engineering design is not all about cognitive skills; it is also about acquiring the right attitudes. Although it is not difficult to illustrate the need for such skills in class, their assessment is more challenging and requires special rubrics. Mourtos (2010) presents an example of a set of rubrics developed to assess open-ended problem solving skills.

A View for Total Engineering

Design engineers must be generalists and acquire a basic understanding of a variety of subjects, from within as well as outside their major – in fact, even from outside of engineering – to develop a view for total engineering. This need has been expressed in three desired attributes for a global engineer (The Boeing Company & Rensselaer Polytechnic Institute, 1997):

- *A good understanding of the design and manufacturing process (i.e., understands engineering and industrial perspective)*
- *A multidisciplinary, systems perspective, along with a product focus*
- *An awareness of the boundaries of one's knowledge, along with an appreciation for other areas of knowledge and their interrelatedness with one's own expertise*

For example, an aircraft designer must have a good understanding of the basic aeronautical engineering disciplines: aerodynamics, propulsion, structures and materials, stability and control, performance, weight and balance. In addition, he/she must develop an understanding of how each part is manufactured and how its design and manufacturing affects the acquisition and operation cost of the airplane.

The example illustrates the multidisciplinary nature of engineering design. Clearly, being an expert in one of the fields involved and inadequate in one or more of the rest, will not work well for a design engineer. Furthermore, engineers must take into consideration a variety of constraints when they design a new product. Some of these constraints are technical; some are non-technical. This expectation is stated in Outcome 3c of ABET EC 2000 (Engineering Accreditation Commission):

"Engineering graduates must have an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability."

The importance of taking into consideration non-technical constraints (e.g., social, political, ethical, safety) is further reinforced in other ABET outcomes as well, where engineering graduates are expected to have:

"3f: an understanding of professional and ethical responsibility.

3h: the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context.

3j: a knowledge of contemporary issues"

- *A basic understanding of the context in which engineering is practiced, including: customer and societal needs and concerns, economics and finance, the environment and its protection, the history of technology and society*
- *High ethical standards (honesty, sense of personal and social responsibility, fairness, etc.)*

In summary, the design engineer must develop an aptitude for systems thinking and maintain sight of the big picture, which is often influenced by technical as well as non-technical factors. Clearly, it is very difficult to quantify a set of specific skills to describe the ideal design engineer. Nevertheless, in an effort to facilitate the teaching and assessment of these design skills, the BSAE Program at SJSU adapted the following set of performance criteria:

Aerospace engineering graduates must be able to:

a. Research, evaluate, and compare aerospace vehicles designed for similar missions.
b. Follow a prescribed process to develop the conceptual / preliminary design of an aerospace vehicle.
c. Develop economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability constraints and ensure that the vehicle they design meets these constraints.
d. Select an appropriate configuration for an aerospace vehicle with a specified mission.
e. Develop and compare alternative configurations for an aerospace vehicle, considering trade-offs and appropriate figures of merit.
f. Apply aerospace engineering principles (ex. aerodynamics, structures, flight mechanics, propulsion, stability and control) to design the various vehicle subsystems.
g. Develop final specifications for an aerospace vehicle.

Ability to Use Design Tools

Freehand Drawing and Visualization

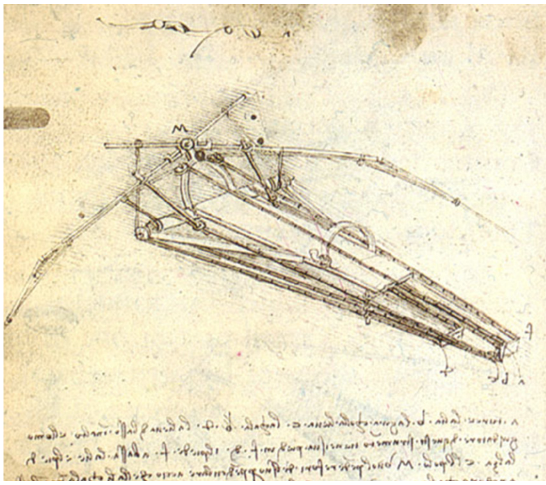
Drawing is the ability to translate a mental image into a visually recognizable form. Eventually any design drawing is rendered as a Computer–Aided Drawing (CAD) with the help of appropriate software. However, CAD is not the best medium when a creative design engineer wants to convey an idea of "how things work" to nontechnical people. Freehand pictorial drawing is most easily and universally understood. Furthermore, a freehand drawing can be a very effective and quick way to communicate ideas in three-dimensions when concepts evolve quickly, as is the case during the early stages of design (e.g., brainstorming), at which point it is not worth investing time and effort in a CAD.

Leonardo da Vinci (1452 – 1519) was one of the earliest engineers who demonstrated mastery in freehand drawing, making it possible for us today to visualize how his inventions worked and appreciate his genius (Figure 3). Freehand drawing is a right-brain activity because it is free of technical symbols and it is closely associated with our ability to visualize things in three dimensions, an indispensable design skill.

Computer–Aided Drawing and Computer–Aided Design

Unlike freehand drawing with its artistic flavor, engineering drawing is a precise discipline based

Figure 3. Design for flying by Leonardo da Vinci (the drawings of Leonardo da Vinci)



on the principles of orthographic projection. In contrast to freehand drawing, engineering drawing emphasizes accuracy, something that has been greatly enhanced by the use of modern computers and graphic capabilities. Today a CAD is much more than a computer generated engineering drawing; it involves an extensive database detailing the attributes of an object and allows it to be rotated, sectioned, and viewed from any angle. This capability is indispensable in the design of complex engineering equipment, such as an airplane, because engineers can now superposition the various subsystems and immediately see potential conflicts. CAD has led to Computer-Aided Manufacturing (CAM), where the machines that manufacture the various components receive their operating instructions directly from the database in the computer.

Kinematics

A design engineer needs skills in kinematics since the various parts of an engineering product move, rotate and may also expand / retract or fold. An understanding of kinematics (e.g., selecting the proper mechanism and visualizing its operation) allows the design engineer to evaluate what will work and what will not work. For example, in the design of an airplane landing gear, the designer must be able to visualize how the gear will fold and retract in its proper space and make sure that it will not conflict with other components in the process.

The skills described in this section fall under Outcome 3k of ABET EC 2000, which states that *engineering graduates must have an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.*

Interpersonal, Communication, and Team Skills

Interpersonal and Team Skills

Archimedes designed his screw pump (Wikipedia, 2007) alone. This was not uncommon in the ancient world. Similarly, Leonardo da Vinci designed his engineering devices, such as the one shown in Figure 3, alone. Today, working alone to design an engineering product is, for the most part, a thing of the past unless, of course, the product is a very simple one. The complexity of modern engineering products requires engineers to work in teams; in fact, sometimes several teams must work together. For example, in the design of a new transport, it is typical to have a team of engineers for each of the disciplines mentioned above (aerodynamics, controls, manufacturing, etc.). These teams work closely together to meet the same set of mission and airworthiness requirements, while at the same time making sure there are no conflicts between the various airplane sub-systems.

Hence, although earlier we expressed the need for design engineers to be generalists, so they can appreciate the multidisciplinary requirements that come into play in the design of a new product, it is not possible for an individual to have enough expertise in each and every one of the technical areas to adequately perform the detail design of all the subsystems, not to mention the analysis of the impact of a new product in a global, economic, environmental, and societal context.

Outcome 3d of ABET EC 2000 states that *engineering graduates must have an ability to function on multidisciplinary teams*. In today's multicultural world, this outcome also implies an ability to collaborate with people from different cultures, abilities, and backgrounds. This is further elaborated in the following four desired attributes for a global engineer (The Boeing Company & Rensselaer Polytechnic Institute, 1997):

- *An awareness of and strong appreciation for other cultures and their diversity, their distinctiveness, and their inherent value.*
- *A strong commitment to team work, including extensive experience with and understanding of team dynamics.*
- *An ability to think both critically and creatively, in both independent and cooperative modes.*
- *An ability to impart knowledge to others.*

The following performance criteria have been chosen to assess this outcome in the BSAE Program at SJSU:

Students working in teams are expected to:

- Be committed to the team and the project; be dependable, faithful, and reliable. Attend all meetings, arrive on time or early, and come prepared and ready to work.
- Exhibit leadership by taking initiative, making suggestions, providing focus. Be creative, bring energy and excitement to the team, and have a "can do" attitude; spark creativity in others.
- Gladly accept responsibility for work and get it done; exhibit a spirit of excellence.
- Demonstrate abilities the team needs and make the most of these abilities by giving fully to the team.
- Communicate clearly with team members when speaking and writing. Understand the direction of the team.
- Bring a positive attitude to the team, encourage others, seek consensus, and bring out the best in others.

Communication Skills

Design requires clear and effective communication not only between team members, but also between the team and third parties (management, customers, etc.). Communication usually takes two forms, oral and written and can be informal, such as between team members or formal, such as when the team presents information to third parties. All four types are crucial for the success of a project. Needless to say, good verbal communication requires not only ability to express one's ideas clearly but also the ability to listen carefully and understand ideas and concerns expressed by others. The need to communicate effectively is outlined in Outcome 3g of ABET EC 2000. In the BSAE Program at SJSU the following performance criteria were selected to express the skills embedded in this outcome:

Ability to:

a. Produce well-organized reports, following guidelines.
b. Use clear, correct language and terminology while describing experiments, projects or solutions to engineering problems.
c. Describe accurately in a few paragraphs a project / experiment performed, the procedure used, and the most important results (abstracts, summaries).
d. Use appropriate graphs and tables following published engineering standards to present results.

It is interesting to note that the desired attribute for a global engineer relating to communication skills, includes *listening* but also *graphic* skills as part of the list (The Boeing Company & Rensselaer Polytechnic Institute, 1997):

"Good communication skills, including written, verbal, graphic, and listening."

Although graphic skills were discussed earlier in the topic of freehand drawing and CAD, the term graphic here includes the ability to prepare engineering graphs that illustrate for example, parametric studies pertinent to a particular design. One thing that becomes obvious in this discussion is that the skills and attributes necessary for competent engineering design are so integrated that in some cases it is not even possible to draw clear distinctive lines between them.

CURRICULUM AND COURSE DESIGN FOR TEACHING ENGINEERING DESIGN SKILLS

Like any set of skills, design skills must be introduced early in the curriculum, practiced often, and culminate in a realistic design experience if students are to achieve the level of mastery prescribed in ABET EC 2000 and expected in industry. The following subsections describe how design skills are introduced at the freshman level, dispersed throughout the BSAE curriculum, and culminate in a senior design capstone sequence. The Project-Based Learning (PBL) pedagogical model is used in all the courses where design is taught and students work in teams for all design projects. Non-traditional ways of assessing design skills are also discussed.

First-Year Design

At SJSU engineering design is first taught in our Introduction to Engineering course (E10). E10 is a one-semester, two-hour lecture/three-hour laboratory course for freshmen, required by all engineering majors. Engineering design is taught through hands-on projects (PBL) as well as through case studies in engineering failures, which also bring up the subject of engineering ethics. For each project, students work in teams to research, brainstorm, design, build, test, and finally demonstrate a device in class (Mourtos & Furman, 2002). Typically, students participate in two or three projects during the semester. This course design followed well-established research, which shows that first-year design courses help attract and retain engineering students (Ercolano, 1996).

E10 students report significant gains in their understanding of design and ethics, design report writing and briefing skills (Mourtos & Furman, 2002). They report slightly lower gains in open-ended problem solving skills, including estimation and mathematical modeling. On the other hand, they report low gains in team skills. This was due to the fact that team skills were not taught explicitly at the time of the assessment. Despite a significant amount of time spent working in teams, students needed more guidance and coaching on skills like conflict resolution, task delegation, decision making, etc.). These skills are now taught more explicitly.

In addition to student self-reporting, authentic assessment data from course instructors show that engineering freshmen perform fairly well in their design assignments.

Design Globally Dispersed-Teaching and Assessment of Open-Ended Problem Solving Skills

In the BSAE Program design is dispersed throughout the curriculum, so students have an opportunity to practice design in a variety of subjects. Student design practice begins with open-ended problems to help them develop the related skills and attributes described earlier. For example, to help students develop:

a. A habit of doing research before attempting to solve a problem: an extensive literature review is required for all open-ended problems and design projects.

b. Competency in the use of a process, as well as specific tactics and heuristics to solve a problem: a problem-solving methodology is taught and required to use in the solution of all open-ended problems.

c. An ability to monitor their progress following a problem-solving process: students write a reflection on the effectiveness of their problem-solving process and identify their strengths and weaknesses.

d. A value system in which accuracy is more important than speed: students are given sufficient time to tackle problems, whether in class (exams) or outside of class, and their grading depends heavily on the accuracy of their calculations.

e. A habit of writing down ideas and creating sketches, charts, and figures while solving a problem: students are graded not only on their final answer but also on how well they integrate such features in their solution of problems.

f. An organized and systematic way of approaching problems: students are expected to document in their solutions every step of the problem-solving methodology they are required to follow.

g. An open mindedness and flexibility when solving problems: students are required to consider, analyze, discuss, and present multiple approaches and solutions to a problem.

h. A risk-taking attitude when solving problems: innovative approaches are encouraged; students are not penalized for presenting such solutions, even when the final outcome is not the best.

i. An ability to use an overall approach that emphasizes fundamentals rather than combining memorized solutions as well as an ability to cope with ambiguity and manage the stress: open-ended problems are practiced in all upper division courses.

Design was originally introduced through projects in several junior level aerospace engineering courses. For example, in aerodynamics (AE162), students designed an airfoil for an ultralight aircraft and a wing for a high subsonic transport, both of which had to meet very specific requirements. Similarly, in propulsion (AE167) students designed a compressor and a turbine and they subsequently matched them for placement in a jet engine with specific thrust requirements.

In an effort to address the compartmentalization of traditional engineering curricula this approach was modified in 2005. In each of the junior fall and spring semesters, students now define their own design project that involves applications from at least two courses, taken concurrently in the particular semester (Mourtos, Papadopoulos, & Agrawal, 2006). For example, one project involved the design of a ramjet inlet and required integration of compressible flow (AE164) and propulsion principles (AE167). Another, more ambitious project involved the design of a flexible wing for high maneuverability and required integration of principles from aerospace structures (AE114), aerodynamics (AE162), flight mechanics (AE165), and computational fluid dynamics (AE169).

This project-based integration of the curriculum offers students an opportunity to appreciate the integrative nature of aerospace engineering design on a smaller scale, before they delve into a much more demanding senior design experience.

Senior Design Capstone Experience

In their senior year, aerospace engineering students may specialize in aircraft (AE171A&B) or spacecraft (AE172A&B) design. Both course sequences involve the conceptual and preliminary design of an aerospace vehicle. Depending on the project, the experience may also include the detail design and manufacturing of the vehicle. Although only one of these course sequences is required, a few students choose to take both in lieu of technical electives.

Teaching and Assessment of Team Skills

As anyone who has ever worked in a team knows, team skills are not acquired automatically simply by working in a team; they need to be taught explicitly, practiced regularly, and assessed periodically, just like any other set of skills. Although team skills are now taught in E10 and assessed in every course that involves a team project or experiment, it is in the senior design course sequence that these skills are formally taught and assessed. As the course meets once a week for two and a half hours, the first 15 to 30 minutes are dedicated to building an understanding of how effective teams work. At the beginning of the year, after teams are formed, students engage in various team-building activities. Lessons from these activities are discussed in class. Subsequently, in each class meeting students present and discuss one of the 17 laws of teamwork (Maxwell, 2001). Finally, at the end of each semester students submit a team member report card, in which they evaluate the performance of their teammates as well as their own, using the performance criteria for effective teamwork defined earlier and which are shown also in Table 1. These peer reviews are taken into consideration when assigning individual course grades.

Table 1. Team member report card

Project Title:					
	Criteria	Filled out by:			
		Member 2	Member 3	Member 4	Self
1	*Quality of Technical Work:* Work is correct, clear, complete, and relevant to the problem. Equations, graphs, and notes are clear and intelligible.				
2	*Commitment to Team / Project:* Attends all meetings. Arrives on time or early. Prepared. Ready to work. Dependable, faithful, reliable.				
3	*Leadership:* Takes initiative, makes suggestions, provides focus. Creative. Brings energy and excitement to the team. Has a "can do" attitude. Sparks creativity in others.				
4	*Responsibility:* Gladly accepts work and gets it done. Spirit of excellence.				
5	Has abilities the team needs. Makes the most of these abilities. Gives fully, doesn't hold back.				
6	Communication: Communicates clearly when he/she speaks and when she/he writes. Understands the team's direction.				
7	*Personality:* Positive attitude, encourages others, seeks consensus. Brings out the best in others.				
	Average score				

Grading scale:

5 – Always, 4 – Most of the time, 3 – Sometimes, 2 – Rarely, 1 – Never

Keep in mind that if you award high scores to everyone, regardless of their contribution, team members who have worked unduly hard or provided extraordinary leadership will go unrecognized, as will those at the other end of the scale who need your corrective feedback.

Please write below and on the back of this form one (**minimum**) or more paragraphs about the work of **each member** of your team, including your own. These narratives should amplify the ratings you gave in the table, by (a) identifying the strengths and weaknesses of each individual and (b) suggesting ways in which his / her work can be more effective. Also, evaluate the team as a whole. Feel free to attach additional pages.

Assessment of Total Engineering Skills

Many student teams choose to participate in the SAE (Society for Automotive Engineering) Aero-Design or the AIAA (American Institute for Aeronautics and Astronautics) Design/Build/Fly competitions. In addition to the conceptual and preliminary design, these teams carry out the detail design of their airplane, which they proceed to build and test. Clearly, these competitions give students an opportunity to go beyond a design on paper and experience challenges related to manufacturability and cost. Often engineering professionals from the aerospace industry mentor students in their designs. Participation in design competitions offers unique learning experiences through interactions with students, faculty, and engineers from educational institutions and companies around the country (US) and the world. Both the SAE and the AIAA competitions attract student teams from universities around the world. Furthermore, it provides unique opportunities for authentic assessment of student design skills by engineering professionals. In addition to the engagement factor, which in itself enhances the students' learning experience in engineering design (Mourtos, 2003), the flight competition itself provides the ultimate test for their designs.

Assessment of Technical Communication Skills

Although students must pass a technical writing course (E100W) and have several design and lab reports evaluated in previous courses, it is again the senior design capstone experience that offers opportunities for more realistic assessment of technical communication skills. For example, students who participate in design competitions have their design reports and drawings evaluated by a team of professional engineers, from whom they receive a score sheet and written feedback. Teams also present their design orally and receive a separate evaluation of their presentation. This kind of feedback naturally adds to any comments given by the course instructor throughout the year. In fact, in many cases it carries a greater weight.

In addition to participating in design competitions, students are encouraged to submit and present papers to conferences (e.g., Johnson et al., 2009; Casas et al., 2008). Whether a student conference or a professional conference, participation provides similar benefits in terms of evaluating student written and oral communication skills.

Safety, Ethics, and Liability Issues

Safety, ethics, and liability issues are addressed in the course through aerospace case studies involving accidents. Students research background information for each case, make a class presentation, and argue about the various issues in class. A written report is also required. Students in general engage in these discussions and perform fairly well in their written assignments not only because safety, ethics, and liability provide an interesting dimension to aerospace vehicle design but also because these assignments are the only ones addressing ABET Outcome 3f in the BSAE Curriculum, and as such, they have been designated as "gateway" assignments. Hence, students must receive a score of 70% or better in these assignments to pass the course, regardless of their performance in the technical aspects of their design.

Economic, Environmental, Societal, and Global Impact

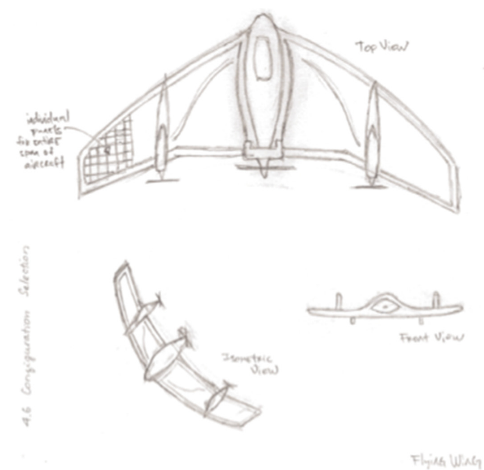
Students discuss in one of their reports the impact of their designs in an economic, environmental, societal, and global context. For example, a team that designed a solar-powered UAV performed a simple analysis on the environmental impact of their airplane by estimating the emissions from a small internal combustion engine with comparable power. They also discussed operating cost taking into consideration the replacement cost of their expensive solar panels every time their UAV crashed. On the other hand, it is not always possible to find interesting and realistic social, political and other types of constraints for all airplanes that students choose to design. Nevertheless, it is important that students develop at least a basic understanding of such issues as well as ways to properly research them before attempting to address them.

To develop such an understanding of these issues as they relate to aircraft design, students perform an additional individual assignment by selecting and researching a topic of interest to them. For example, two very interesting topics selected by students were the impact of airplanes on cultural integration and the contribution of jet aircraft contrails on global warming. Students are required to find at least five references related to their topic, at least two of which must be technical journal articles, conference papers or technical reports. For the rest of their references students may use newspaper or magazine articles and the worldwide web. Students study these references and prepare a two-page paper summarizing the key points of their research and a ten-minute presentation for our class. In their presentation students must include two key questions related to their issue, as a way to facilitate class discussion.

Graphic Communication Skills

To introduce students to freehand drawing a collaboration has been established with the SJSU School of Art and Design. A team of students from the graduate class Artists Teaching Art Seminar (Art 276) visits the aircraft design class to offer a three-hour workshop on freehand drawing, which includes contour drawing, gesture drawing, and perspective. Both groups of students have been very positive about their experience; the art students because they are given an opportunity to practice their teaching skills in a realistic setting; the aircraft design students because they get an opportunity to express themselves creatively

Figure 4. Example of a free-hand drawing in the early design stages



within the context of a very demanding engineering course. An example of a free-hand drawing illustrating a possible configuration for a small solar-powered UAV is shown in Figure 4.

Engineering students tend to be very capable with computer programs, including those used in design. For example, a student produced an artist's concept of his proposed very large, luxury airship, as a way of helping his audience visualize the level of comfort and luxury afforded in this kind of vehicle and provide a contrast with the interior one finds in most airlines today (Figure 5).

Naturally, three-view CAD drawings are expected from students in all final design reports. Students are introduced to CAD early in their curriculum with a required freshman-level course in Design & Graphics (ME20). In addition, Computer Aided Design (ME165) is a popular technical elective for many students.

REQUIRED SKILLS FOR FACULTY WHO TEACH ENGINEERING DESIGN

An additional challenge in teaching design is the competence level, as far as design skills are concerned, of the faculty who teach design courses.

Figure 5. Example of an artist's concept drawing for the interior of a very large, luxury airship

To provide a thorough analysis of this issue is beyond the scope of this article, however, it is worth mentioning two very distinct reasons, which contribute to this challenge:

a. A successful completion of a Ph.D. degree, required for a faculty position at most engineering schools, entails primarily development of analytical (left brain), research skills. On the other hand, as we have seen, design requires both analytical and creative skills.
b. To earn tenure and promotion in an academic setting engineering faculty are required to perform research, publish in refereed journals, and seek external funding. To maximize their chances for success under this kind of pressure, engineering faculty continue the same line of research they did in graduate school. After all, the venues available for publishing design work or seeking funding to do such work are limited compared with traditional areas of engineering research.

Hence, faculty members who are asked to teach a design course, often find themselves unprepared. One way to address this deficiency is to require engineering faculty to undergo some training in engineering design before teaching a design course. There are many workshops on design for faculty members as well as for engineers who work in industry, sponsored by professional societies, universities, and engineering companies. Professional societies also offer summer fellowships for engineering faculty willing to spend a summer in industry working alongside design engineers. Another way to address this issue is to hire adjunct faculty with current design experience from industry to teach design courses. This solution, however, poses its own problems.

a. While some engineering schools are strategically located in areas where adjunct faculty with design experience are available, not every engineering school is blessed with proximity to engineering companies that may provide such faculty. This issue can be addressed in creative ways. For example, to accommodate an adjunct faculty member who teaches a design course at SJSU, a blended course has been scheduled: traditional (face-to-face) and online. The instructor flies from another state every other week and spends three hours with the students. In between, the course is conducted online using appropriate software.
b. Teaching any subject including design requires not only expertise in the subject matter but also appropriate pedagogical

knowledge (Mourtos, 2007). Unfortunately, most engineering faculty do not possess such knowledge, as it is not a requirement in their job description. This is true for full-time as well as part-time faculty. Our experience at SJSU has shown that both full-time and adjunct faculty have opportunities to develop pedagogical knowledge through experience and reflection by teaching a variety of courses over time as well as through optional pedagogical training available at most universities. As a result, some – certainly not all – of the faculty do develop appropriate pedagogical content knowledge over time and become effective teachers.

CONCLUSION

An attempt has been made to provide a comprehensive list of skills, technical and non-technical, for design engineers. These skills include analytical, open-ended problem solving, a view for total engineering, interpersonal and team skills, communication skills, as well as fluency with modern tools and techniques used in engineering design. In addition to these skills, design engineers must develop certain attributes, such as curiosity to learn new things and explore new ideas, self-confidence in making design decisions, taking risks by trying new concepts, thinking out-of-the-box, and persistence to keep trying when things don't work.

The paper presented course and curriculum design from the BSAE Program at SJSU that addresses these skills and attributes and touched briefly on the challenge of engineering faculty competence in design skills and pedagogy. Some of the elements in this curriculum were introduced several years ago, have been assessed extensively and indicate that students indeed acquire an adequate level of competence in some of these skills. Some of these elements, such as the teaching of freehand drawing through the collaboration with the College of Arts and Design, were introduced only recently and have not yet been assessed. In any case, the attributes of a design engineer, as described above, are difficult to measure and will require the development of special rubrics.

REFERENCES

Bloom, B. S. (1984). *Taxonomy of educational objectives; Handbook 1: Cognitive domain*. Reading, MA: Addison-Wesley.

Bloom, B. S., Karthwohl, D. R., & Massia, B. S. (1984). *Taxonomy of educational objectives; Handbook 2: Affective domain*. Reading, MA: Addison-Wesley.

Casas, L. E., Hall, J. M., Montgomery, S. A., Patel, H. G., Samra, S. S., Si Tou, J., et al. (2008). Preliminary design and CFD analysis of a fire surveillance unmanned aerial vehicle. In *Proceedings of the Thermal-Fluids Analysis Workshop*.

Curry, D. T. (1991). Engineering schools under fire. *Machine Design*, *63*(10), 50.

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education*, *94*(1), 103–120.

Engineering Accreditation Commission & Accreditation Board for Engineering and Technology. (2009). *Criteria for accrediting engineering programs, effective for evaluations during the 2010-2011 cycle*. Retrieved August 20, 2010 from http://www.abet.org/forms.shtml

Ercolano, V. (1996). Freshmen: These first-year design courses help attract and retain engineering students. *ASEE Prism*, 21-25.

Flightsim Aviation Zone. (2010). *Federal Aviation Regulations, Part 25 – Airworthiness Standards, Transport Category Airplanes*. Retrieved April 18, 2011 from http://www.flightsimaviation.com/data/FARS/part_25.html

Flightsim Aviation Zone. (2010). *Federal Aviation Regulations, Part 36 – Noise Standards*. Retrieved April 30, 2011 from http://www.flightsimaviation.com/data/FARS/part_36.html

Johnson, K. T., Sullivan, M. R., Sutton, J. E., & Mourtos, N. J. (2009). Design of a skydiving glider. In *Proceedings of the Aerospace Engineering Systems Workshop*.

Maxwell, J. C. (2001). *The 17 indisputable laws of teamwork: Embrace them and empower your team*. Nashville, TN: Thomas Nelson.

Mourtos, N. J. (2003). From learning to talk to learning engineering; Drawing connections across the disciplines. *World Transactions on Engineering & Technology Education*, *2*(2), 195–204.

Mourtos, N. J. (2007). *Course design: A 21st century challenge* (pp. 1–4). San Jose, CA: Center for Faculty Development and Support, San Jose State University.

Mourtos, N. J. (2010). Challenges students face when solving open - ended problems. *International Journal of Engineering Education*, *26*(4).

Mourtos, N. J., DeJong-Okamoto, N., & Rhee, J. (2004). Open-ended problem-solving skills in thermal-fluids engineering. *Global Journal of Engineering Education*, *8*(2), 189–199.

Mourtos, N. J., & Furman, B. J. (2002). Assessing the effectiveness of an introductory engineering course for freshmen. In *Proceedings of the 32nd IEEE/ASEE Frontiers in Education Conference.*

Mourtos, N. J., Papadopoulos, P., & Agrawal, P. (2006). A flexible, problem-based, integrated aerospace engineering curriculum. In *Proceedings of the 36th IEEE/ASEE Frontiers in Education Conference*.

Nicolai, L., & Pinson, J. (1988). *Aircraft Design Short Course*. Dayton, OH: Bergamo Center.

Nicolai, L. M. (1998). Viewpoint: An industry view of engineering design education. *International Journal of Engineering Education*, *14*(1), 7–13.

Petroski, H. (2000). Back to the future. *ASEE Prism*, 31-32.

Pink, D. H. (2005). *A whole new mind: Why the right-brainers will rule the future*. New York, NY: Riverhead Books.

Reuteler, D. (2010). *The drawings of Leonardo da Vinci*. Retrieved August 20, 2010, from http://www.drawingsofleonardo.org/

Seely, B. E. (1999). The other re-engineering of engineering education, 1900-1965. *Journal of Engineering Education*, 285-294.

The Boeing Company & Rensselaer Polytechnic Institute. (1997). *A Manifesto for Global Engineering Education: Summary Report of the Engineering Futures Conference*. Seattle, WA.

Wikipedia. (2007). *Archimedes' screw*. Retrieved August 18, 2010 from http://en.wikipedia.org/wiki/Archimedes%27_screw

Wikipedia. (2011). *Hughes H-4 Hercules*. Retrieved April 18, 2011, from http://en.wikipedia.org/wiki/Hughes_H-4_Hercules

Woods, D. R., Hrymak, A. N., Marshall, R. R., Wood, P. E., Crowe, C. M., & Hoffman, T. W. (1997). Developing problem-solving skills: The McMaster problem-solving program. *Journal of Engineering Education*, *86*(2), 75–91.

This work was previously published in the International Journal of Quality Assurance in Engineering and Technology Education (IJQAETE), Volume 2, Issue 1, edited by Arun Patil, pp. 14-30, copyright 2012 by IGI Publishing (an imprint of IGI Global).

Chapter 2
Why Get Your Engineering Programme Accredited?

Peter Goodhew
University of Liverpool, UK

ABSTRACT

In many countries engineering degree programmes can be submitted for accreditation by a professional body and/or graduate engineers can be certified or registered. Where this is available most academic institutions feel that they must offer accredited engineering programmes. The author suggests that these processes are at best ineffective (they do not achieve their aims) and at worst they are destructive of creativity, innovation and confidence in the academic community. The author argues that such processes (including any internal certification within the Conceive-Design-Implement-Operate, i.e., CDIO Initiative) should be abandoned completely. The author proposes alternative ways of maintaining the quality of engineering design and manufacture, which place the responsibility where it properly lies – with the manufacturer or contractor. This is a polemic piece, not a referenced review of accreditation.

INTRODUCTION

In many countries undergraduate engineering programmes can be submitted to a national body for *accreditation*. Graduates from accredited programmes are eligible, often with an additional requirement for relevant work experience, for registration as a professional engineer. In the UK this accreditation is overseen by the Engineering Council via UK-Spec. and opens the way to C.Eng, I.Eng or Eng Tech qualifications. In the USA ABET serves a similar function, while in Australia the appropriate body is Engineers Australia. In all cases the programme, its students, and sometimes its graduates, are scrutinised by a committee of professional engineers before accreditation is awarded for a fixed period such as five years. The accreditation process involves substantial paperwork and usually a one or two day visitation, so is quite costly both for the educational institution and the professional body. I argue in this article that this considerable effort does not represent good value for money and in some cases may have a negative effect on the quality of engineering education.

DOI: 10.4018/978-1-4666-1945-6.ch002

THE CASE AGAINST ACCREDITATION

Did the accreditation of professional engineering programmes prevent the disastrous crash of the Airbus 330, flight AF 447, in June 2009? Equally, is it responsible for the fact that the Eiffel tower has remained standing for 120 years? Or that my iPhone is so brilliant? No, no and no. So what is accreditation supposed to be for? At the highest level I presume that the intention is to ensure and enhance the quality and safety of engineered products throughout the world. At a more mundane (and self-interested) national level it might be intended to enable the world-wide transferability, and thus profitability, of a nation's engineering industry by ensuring the international credibility and employability of its engineers.

These seem to be laudable objectives, but delivery of them is several steps away from the accreditation of university programmes. The logic is presumably that the employers of professional engineers must have confidence, via external testimony, in their skills and their fitness to practice. This confidence is engendered by their status as professional (*chartered* in UK parlance, *registered* in other jurisdictions) engineers, part of the qualification for which is that, at some time in the past, they graduated from an *accredited* degree programme. These engineers also have to demonstrate some appropriate experience in employment and the membership of a professional body.

I find the whole system of accreditation unsatisfactory in two ways: It does not deliver the intended outcome (and so is ineffectual) and, additionally, it can damage our education system and thus our students and graduates.

First, the charge that it is ineffectual: Engineered products are conceived, designed, made and operated (CDIO-ed) by engineers employed by large or small companies. Some, but certainly not all, of these engineers may be chartered. They will usually have earned their chartered status by virtue of the work undertaken in their first few years of employment, backed up by the degree they were awarded several years ago. Since receiving their chartered status they will have been encouraged to undertake continuous professional development, but this will not have been checked. A fifty-year-old chartered engineer is thus operating on the basis of a validation process twenty years ago and a degree awarded about 25 to 30 years ago. The accreditation of this degree, so long ago, has almost no relevance for the engineering practices in use today. Indeed if the degree was typical of those awarded 25 years ago it will have contained a significant amount of engineering science and very few tests of engineering aptitude or attitude (which is of course why we have the CDIO *movement*). The fitness to practice of an individual engineer will in reality depend on what they have done, seen and learned during their working life, which is almost independent of the content of their first degree. Indeed the technical content of a degree in one engineering discipline may have almost no overlap with the content of another engineering discipline so it is hard to argue that subject content has anything to do with being, or thinking like, an engineer.

Furthermore an engineer employed today may be working in an area unrelated to their original area of study. This is very likely for bioengineers, nanoengineers, environmental engineers, nuclear engineers and others working in interdisciplinary areas. Their original degree would either have been un-accredited or the accreditation would relate to a different disciplinary area. How can this in any way validate or assure the quality of their current work?

A third issue is the effectiveness of the quality assurance provided by chartered status. I have already asserted that there are almost no checks on the continued professional development of chartered engineers, but equally there are almost no cases of the de-registration of rogue chartered engineers (and even if there were, they would certainly – like doctors – be de-registered after they had committed a grave misjudgment or offence, not before).

So the accreditation of programmes is certainly ineffectual, but it is also damaging to the education process. University departments of Engineering spend a great deal of time preparing for accreditation visits, and tuning their degree programmes to fit the perceived requirements of their professional bodies. They do this not to improve their programmes (most programme leaders do not believe that the comments of accreditors will achieve this) but because of the fear that they will no longer be able to compete in the marketplace for students if they are not accredited. This fear is probably misplaced, but no department has the courage to put it to the test. Accreditation panels almost always feel that they should make some critical (framed as *helpful*) comments but these usually reflect the prejudices of individual panel members, who are rarely experts in higher education and frequently elderly and tending to be out of date. (I have resolved never to accept another invitation to sit on an accreditation panel now I have reached 65.) The damage to the system is that the threat of accreditation makes our engineering departments more conservative, less willing to change or innovate, as well as taking time and money which would be better spent on the education of their students. It also reinforces (unhelpfully) the audit culture which has over-run our universities in the last twenty years (at least in the UK).

It would be unreasonable to criticise the existing system of accreditation without making some attempt to suggest what might replace it to provide the assurance of quality demanded by society. My suggestion is that the responsibility for the safety and quality of products (from multi-billion tunnels to five-penny toys) should remain where it legally is – with the manufacturer or major contractor. These businesses should assure themselves that their workers are appropriately skilled and work to appropriate safety and ethical standards. To achieve this they might need to strengthen their recruitment procedures to include a real assessment of candidates' current abilities and skill sets. They would also want, as many do, to ensure periodically that their employees are up to date. They might wish to buy in the necessary training expertise, perhaps even from a local university, but they will not be much helped by a past accreditation. The proof of the quality of training, and of initial education, will be demonstrated by the performance of the employee – supervised and checked by experienced colleagues – not by their possession of a yellowing piece of paper.

I notice that I have not mentioned professional bodies. What might their role be? Certainly not as accreditors, but perhaps as honest brokers between employers and trainers and educators, or as forums for discussion (but not regulation) of best practice. In which case perhaps there should be an upper age limit for service on any committee or as an officer – shall we say 50 – and those in their dotage (like me) should only speak when asked.

CONCLUSION

The arguments I have advanced here also apply to the certification of undergraduate programmes as, *CDIO-compliant*. Such a scheme would cost effort (and almost certainly money) to implement, it would cost even more to police (so this would be unlikely to happen) and would still offer no assurance of the quality of an engineering graduate. A further particular argument which applies to CDIO members is that (unlike many other engineering teaching departments) they have already shown their commitment to improving engineering education and are thus the least likely programmes to need the additional discipline offered by certification process. So I strongly suggest that we do not bother.

This work was previously published in the International Journal of Quality Assurance in Engineering and Technology Education (IJQAETE), Volume 2, Issue 2, edited by Arun Patil, pp. 93-95, copyright 2012 by IGI Publishing (an imprint of IGI Global).

Chapter 3
Quality and Environmental Management Systems in the Fashion Supply Chain

Chris K. Y. Lo
The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

Consumers and stakeholders have rising concerns over product quality and environmental issues, and therefore, quality and environmental management have become important topics for today's fashion products manufacturers. This chapter presents some empirical evidence of the adoption of quality management systems (QMS) and environmental management systems (EMS) and their impact on fashion and textiles related firms' supply chain efficiency. Although both management systems are commonly adopted in the manufacturing industries and becoming a passport to business, their actual impacts specifically on the fashion supply chain have not been explored. By investigating the adoption of ISO 9000 (a quality management system) and ISO 14000 (an environmental management system) in the U.S. fashion and textiles firms, we estimate their impact on manufacturers' supply chain performance. Based on 284 publicly listed fashion and textiles manufacturing firms in the U.S., we find that fashion and textiles firms operating cycle time had shortened by 15.12 days in a five-year period. In the cross-sectional analysis, the results show that early adopters of ISO 9000 and high-tech textiles related firms obtained more supply chain benefits. We only find mixed results of the impact of ISO 14000 on supply chain performance.

BACKGROUND

The quality of textiles products at each stage in the fashion supply chain is essential for the success of a fashion product. The quality level delivered to the final customer is the result of quality management practices of each link in the fashion supply chain, thus each actor is responsible for their own quality issues (Romano & Vinelli, 2001). This is because the quality of the final product that reaches the customers is clearly the results of a chain of successive, inter-linked phases: spinning, weaving, apparel and distribution, and thus quality management in supply chain are particularly

DOI: 10.4018/978-1-4666-1945-6.ch003

relevant in the fashion and textiles industries (Romano & Vinelli, 2001).

Quality management is defined as an integrated approach to achieve and sustain high quality output, focusing on the maintenance and continuous improvement of processes and defect prevention at all levels and in all functions of the organization to meet or exceed customer expectation (Flynn, Schroeder, & Sakakibara, 1994). The customer expectations on the quality of a product, however, are not just its physical attributes and workmanship. According to ISO 9000, quality is defined as customer expectations over actual performance (ISO, 2004). Consumers' expectations on fashion products, nowadays, also include its environmental attributes, for instance, use of sustainable materials and control of environmental impacts during the manufacturing processes, etc. Therefore, both quality and environmental management have become important focuses for today's fashion and textiles manufacturers. International buyers of major brands often use quality management system (QMS) and environmental management systems (EMS) as a major tool to select capable fashion and textiles suppliers (Boiral, 2003; Boiral & Sala, 1998), to ensure their products and raw materials could meet customers' expectations on quality and environmental aspects.

To respond to the call for management systems in various industries, International Organization for Standardization (ISO) has developed ISO 9000 in 1987 and ISO 14000 in 1996, which are generic QMS and EMS for worldwide applications. The number of ISO 9000 certified firms has been increasing persistently since its introduction some 20 years ago. According to recent statistics (ISO, 2009), almost one million of firms or business divisions in 175 countries have adopted ISO 9000. In the past five years, almost 800,000 firms or business units have adopted ISO 9000, representing an increase of almost 570%. For ISO 14000, it has been adopted by 188,815 firms or business divisions in 155 countries (ISO, 2009). From 2006 to 2008, almost 60,000 firms or business units have adopted ISO 14000, representing an increase of about 47% (ISO, 2009). Multinational enterprises (MNEs) with operations in more than one country are widely recognized as key agents in the diffusion of ISO certifications across national borders.

The diffusion of ISO 9000 in the fashion and textiles industries is particularly pronounced. In the early 1990s, European Committee for Standardization (Committee European pour Normalization - CEN) developed importing regulations for use by the European Union (EU) countries. CEN requires manufacturing firms that are importing products into the European market to comply with ISO 9000 standard. Import of fashion and textiles products to EU countries are under this regulation. The requirement of ISO 9000 was then followed by major MNEs, which use the ISO-based criteria to certify their own suppliers and have developed their internal quality management systems according to the ISO guidelines (Guler, Guillen, & Macpherson, 2002). Many suppliers to MNEs subsequently required their upstream suppliers or business partners to be ISO certified, leading to the widespread diffusion of the standard in global supply chain.

ISO 14000 follows the global diffusion pattern of ISO 9000, and it has become the most widely adopted EMS in the world (Corbett & Kirsch, 2001). It is a set of management processes and procedures requiring firms to identify, measure, and control their environmental impacts (Bansal & Hunter, 2003). With the aim of improving the environmental performance of a firm, compliance with the standard is audited and certified by an independent, third-party certification body (Jiang & Bansal, 2003). The initial version of ISO 14000 was a consolidation of various elements in BS 7750, a British environmental management standard, and European's Environmental Management and Audit Scheme (EMAS).

Regulations of different countries towards the adoption of ISO 14000 also affect the diffusion of this standard. European countries and some Asian countries, such as Japan (Bansal & Roth, 2000) and Singapore (Quazi, Khoo, Tan, & Wong, 2001), provide favourable legislative environment for firms to adopt EMS, while the U.S. comparatively provides less favourable legislative environment (Kollman & Prakash, 2001, 2002). The regulatory environment within a country affect the costs and perceived benefits of ISO 14000 adoption (Delmas, 2002). Jennings and Zandbergen (1995) maintained that the larger the pressure from the environment, the faster the diffusion of EMS.

Due to globalization, the environmental law and regulations are not just affecting the firms within a particular country, but also affecting firms which import and export goods to other countries. Christmann and Taylor (2001) found that if the polluting firms in developing countries export a large proportion of their output to developed countries, they would be more likely to adopt ISO 14000, despite they may be tempted by lax environmental regulations in developing countries.

Although the original objective of ISO 14000 is quite different from ISO 9000's, they shared the same management framework and both diffuse along the global supply chain. In the literature, some scholars investigated on the interactions between ISO 9000 and ISO 14000. Corbett & Kirsch (2001) found that ISO 9000 appears as an important factor explaining diffusion of ISO 14000, suggesting that the motivation, such as attracting potential customers, behind the two have significantly overlap. Pan (2003) also found that there is a strong linkage between the motivations of implementing ISO 9000 and ISO 14000 and their perceived benefits of adoption. Albuquerque, Bronnenberg, & Corbett (2007) further investigated the global diffusion patterns of ISO 9000 and ISO 14000, and they found that both certifications diffuse across countries primarily by geography. In addition, the adoption experience of ISO 9000 could also help certified firms to effectively implement ISO 14000, as the two systems are very similar in terms of implementation requirements (Poksinska, Dahlgaard, & Eklund, 2003).

LITERATURE REVIEW OF QMS AND EMS ADOPTION IN THE FASHION SUPPLY CHAIN

In fashion and textiles literature, researchers mainly focus on the usefulness of QMS and EMS in supplier selection process. Buyers use management certification (e.g., ISO 9000 and ISO 14000) as an instrument to determine whether the supplier is capable to follow industry's standards (Motwani, Youssef, Kathawala, & Futch, 1999; Teng & Jaramillo, 2005; Thaver & Wilcock, 2006). Teng and Jaramillo (2005) developed a model for evaluation of supplier in the fashion supply chain. They proposed five performance clusters to evaluate textiles supplier performance, which are delivery, flexibility, cost, reliability, and quality. They suggested using QMS certifications as the evidence for quality performance evaluation. The suppliers would receive a higher score in the evaluation model, if they are ISO 9000 certified.

There are only a few anecdote cases that discussed the impact of QMS on the fashion supply chain. Sarkar (1998) found that a textiles mill in India obtained higher customer satisfaction through increased employee involvement and product quality improvement due to ISO 9000 adoption. Romano and Vinelli (2001) conducted a case study of the Marzotto Group, one of the most important Italian textiles and apparel manufactures, about its relationships with both upstream and downstream suppliers. They found that quality management system is the "glue" that makes the supply network to operate as a "whole system". Adanur and Allen (1995) conducted the first industry survey of ISO 9000 in the U.S. textiles industry, and they found that the certified firms experienced decrease in production time and product returns. The certified firms also reported fewer raw material rejections from ISO 9000 certified suppliers. A recent study shows that the adoption of ISO 9000 can improve the adopting firms' supply chain efficiency in the general manufacturing industries in the U.S. (Lo, Yeung, & Cheng, 2009), suggesting that the adoption of QMS might help the certified firms to become a more efficient node in their supply chain.

The implementation of QMS are often under the pressure of major customers, thus small- and medium-sized textiles manufacturers have no choice but to comply with customer requirements on these certification. They might pursuit ISO 9000 without genius knowledge about QMS. Allen and Oakland (1988, 1991a, 1991b) found that small textiles firms lack correct knowledge about QMS compared to large firms, in three survey studies of 183 textiles firms. They concluded that there is a distinct lack of good quality management practices within in the British textiles industry. Fatima and Ahmed (2010) studied the ISO 9000 certified firms in Pakistan's bedwear textiles industry. They found that 60% of the firms offered poor training, 70% had poorly defined quality policy and objectives, and 70% had ineffective internal audit. Their findings show that despite the high adoption rate of ISO 9000 in the industry, there is lack of real implementation of QMS, and ISO 9000 is merely a passport into export markets. Fashion and textiles manufacturers in developing countries have self-regulation pressures to adopt QMS for gaining legitimacy from the MNEs of developed countries (Christmann & Taylor, 2001).

In the operation management (OM) literature, there are some critics on EMS's benefits in firm operations. The critics believe that environmental initiatives often transfer the cost previously bored by the society back to the firms (McGuire, Sundgren, & Schneeweis, 1988). The increased liability and environmental obligations would also lead to a negative impact on firms' operational performance and its flexibility in production (i.e. limited choices of environmental friendly dyes and fabrics, which are often more expensive). Therefore, textiles firms' operations managers would hesitate to adopt EMS in their production. On the other hand, the advocate of EMS believe that the ISO 14000 could improve environmental performance, which eventually improved firms' performance through cost-saving and revenue gain pathways (Klassen & McLaughlin, 1996).

In fact, the perceived benefits of EMS in fashion supply chain are quite direct. The dying process in textiles processing could produce huge amount of emissions that would lead to fine and high restoration cost. The impact of EMS is especially important for the wet processing of natural fibres, which are water, energy and pollution intensive (Ren, 2000). Managing the environmental impact through a systematic process would help the firms to reduce water and energy consumption, as well as to avoid serious pollution incidents. Therefore, EMS adoption is particularly relevant to fashion and textiles related firms. However, compared to ISO 9000, the number of empirical works that focused on the EMS adoption in the fashion supply chain is very limited. There are only a few survey studies discussed about the sustainability of the fashion supply chain.. For example, Fresner (1998) conducted a case study of an Austrian textiles mill and found that the adoption of ISO 9000 helped to reduce solid waste and then pursuit ISO 14000 for further improvement in productivity. Brito et al. (2008) found that the adoption of ISO 14000 in Europe fashion and textiles industries would improve customer services and cost optimization for the adopting firms, and eventually improve the overall performance of the whole supply chain. However, both studies mentioned above did not provide objective evidence of such impact. To explore how QMS and EMS affect fashion and textiles firms' supply chain efficiency, more empirical evidence is needed.

HYPOTHESIS DEVELOPMENT

Supply chain efficiency could refer to lead-time performance, delivery promptness and inventory level (Kojima, Nakashima, & Ohno, 2008). If the members of a fashion supply chain could deliver their products to their down stream customers in shorter period, the performance of the overall supply chain performance is improved. The well-known Supply Chain Operations Reference (SCOR) model suggested that firms' operating cycle time is an important indicator to measure supply chain efficiency (Stewart, 1995, 1997). To measure supply chain efficiency for a particular firm, accounting-based performance indicators, namely 1) inventory days, 2) accounts receivable days, and 3) operating cycle time can be used (Lo, et al., 2009). We discuss how ISO 9000 and ISO14000 could possibly affect the supply chain efficiency of adopting firms as follows.

ISO 9000 requires the adopting firm to ensure that product quality is constantly measured and appropriate corrective actions are taken whenever defects occur. These actions must be undertaken through a well-defined management system that monitors the potential quality problems continuously. Therefore, the defect rate of the fashion products should decrease and defects should be detected and corrected early in the production processes, and less scrap and rework need to be handled in the manufacturing processes (Naveh & Marcus, 2005). Therefore, the overall time required turning raw material into fashion and textiles products that fulfil customer order should be shortened (i.e. lower inventory days).

ISO 14000 adoption could also lead to lower inventory days, as it requires organizations to implement pollution prevention procedures to avoid environmental spills, crises, and liabilities that might incur huge effort in restoration (Bansal & Hunter, 2003; Brio, Fernandez, Junquera, & Vazquez, 2001). Therefore, ISO 14000 certified firms could face less frequent mandatory pollution restoration that could seriously disrupt their operations. ISO 14000 certified firms are also perceived as less risky than their non-certified counterparts, less frequent environmental inspections from customers and regulators are required (Potoski & Prakash, 2005), leading to further shortening of inventory days. Therefore, we hypothesize that the time required to convert raw materials into textiles products (i.e., inventory days) is shorter after ISO 9000 or ISO 14000 implementation. As the initial objectives and its impact on supply chain are different between ISO 9000 and ISO 14000, we develop two parallel hypotheses to estimate their impact on supply chain performance independently.

H1a: *The adoption of QMS leads to lower inventory days.*

H1b: *The adoption of EMS leads to lower inventory days.*

The perceived benefits of ISO 9000 are not just confined to improving product quality, but also enhancing customer services (Buttle, 1997). If ISO 9000 implementation can improve product quality and customer services, the customer orders fulfilment time should be shorter. Moreover, if there were any quality problem of the fashion products, payment would be postponed as defective fashion products are returned to rework. Customers may not pay for the products until the reworked products being delivered and met their quality requirements. As a result, the time between product delivery and customer payment (i.e. accounts receivable days) should be shorter for firms with higher product and service quality for the ISO 9000 certified firms (Lo, et al., 2009).

By taking proactive measures to prevent environmental crises, an ISO 14000 certified firm are able to prevent mistaken use of hazardous materials, and violation of environmental regulatory requirements in the customer's market, which might results in large-scale product recall. Once a product recall is needed, the accounts receivable days will be significantly affected (i.e. longer). Besides, as customers are more favourable toward environmentally friendly products and organizations, ISO 14000 adoption can establish a positive corporate image and trust from customers in the long run (Bansal & Hunter, 2003). Therefore, such firms have the potential to bargain for more favourable payment terms from customers who trusted and loyal to them (i.e., customers are willing to pay earlier). This hypothesis can be tested by measuring the accounts receivable days.

H2a: *The adoption of QMS leads to lower accounts receivable days.*
H2b: *The adoption of EMS leads to lower accounts receivable days.*

In the fashion supply chain, operating cycle time consists of manufacturing time (the time required to turn raw materials into products), delivery time (the time required to deliver products from the manufacturer to customers), and payment fulfilment time (the time required for customers to pay for their accepted products). The total time incurred in the above processes is known as operating cycle or "cash-to-cash cycle" (Eskew & Jensen, 1996). Therefore, we hypothesize that operating cycle time should be shorter after the implementation of ISO 9000 and ISO 14000.

H3a: *The adoption of QMS leads to a shorter operating cycle.*
H3b: *The adoption of EMS leads to a shorter operating cycle.*

METHODOLOGY AND DATA COLLECTION

In this research, we focus on fashion and textiles related firms in the manufacturing industry (SIC code 2000-3999). All firms under the industry that contains keywords such as "Fashion", "Textiles", "Dye", "Apparels" and "Fabrics", were included in our sample. To generate our sample, we identified ISO 9000 and ISO 14000 certified firms and their years of certification from registration data of *Quality Digest* and *Who's Registered,* which are online registration databases. Since each firm could have multiple plants/sites certified, we follow the practice of previous research (e.g., Corbett, Montes-Sancho, & Kirsch, 2005; Naveh & Marcus, 2005) by focusing on the first certification. It is because this is the only time period representing the change from a non-certified to a certified firm status. Additional certifications after the first certification only mean continuous improvements. After compiling the data from the online databases and from Standard and Poor's financial database - COMPUSTAT, we found that 284 ISO 9000 certified and 61 ISO 14000 certified publicly listed fashion or textiles manufacturing firms in the U.S.

We define the year of formal ISO certification as the certification year (year 0). To measure the abnormal change in performance over a long-term period (event window), we should start by defining the base year, when there is no prior impact from the preparation of ISO certification on the sample firms. To pass the certification audit, the average preparation time is 6-18 months prior to registration (Corbett, et al., 2005). Therefore, year -2 should be taken as the base year. As we only focus on the first certification, we can assume that there is no impact from ISO implementation on all the sample firms during the base year (year -2). ISO requires a third-party audit to verify that it has been effectively implemented, so a strong impact of ISO on performance should appear in the certification year when the firm passed the audit. The performance of certified firms should also experience the impact years after ISO certification. Therefore, we set the event period in this research at year -2 as the base year and measure the changes over the next five years (year -1, year 0, year 1, year 2, and year 3).

To estimate the abnormal changes in supply chain performance within the event window, we compare the actual performance with the expected performance of the sample firms, which is based on the changes in performance of control firms (Barber & Lyon, 1996). The selection of control firms should be based on a combination of three criteria: pre-event performance, industry, and firm size (Barber and Lyon, 1996). First, matching on pre-event performance can avoid the mean reversion problem of accounting data and control the impact of other factors on firms' performance (Barber & Lyon, 1996). Second, industry economic status could account for up to 20% of the changes in financial performance (McGahan & Porter, 1997). Moreover, environmental issues and the impact of EMS are industry-specific (Russo & Fouts, 1997), so we must control industry type in matching sample and control firms. Third, previous studies have suggested that operating performance varies by size (e.g., Fama & French, 1995). We thus match sample and control pairs based on the three matching criteria.

We generate the sample-control pairs and regard these pairs as the *performance-industry-matched* group. The first step is to match each sample firm to a portfolio of control firms based on at least a two-digit SIC code and 90%-110% of performance in year -2. In Step 2, if there are some firms have no control firm is matched in step 1, we use at least a one-digit SIC code and 90%-110% of performance to match for control firms. If no control firm is matched in the Step 2, we use only the 90%-110% of performance as the matching criterion.

In our sample, there are 248 ISO 9000 certified firms, of which 58 are ISO 14000 certified. We discard 57 firms that did not have financial information (Operating cycle) in year -2. In the remaining 191 fashion and textiles related firms, 154 firms are matched in Step 1 (80.6%). five firms are matched in Step 2. No firm is matched in Step 3. We matched 159 sample firms with at least one control firm in year -2. The financial information presented in Table 1, the number of sample and control matched pairs gradually decrease from the ending period year 1 to year 3, due to lack of financial information in either sample firms or control firms.

For another group of matching, i.e., *performance-industry-size-matched*, we further control the firm size. We use the 50%-200% range of total assets for controlling the firm size. In other words, we match the sample and control firms with at least a two-digit SIC code, 50-200% firm size (in terms of total assets), and 90%-110% prior certification performance in year -2 (Step 1). In the case where a sample firm cannot match any control firm in Step 1, we relax the industry-matching criterion to at least a one-digit SIC code, while keeping the other matching criteria unchanged (Step 2). If no control firm is matched in Step 2, we use only the 50%-200% of firm size and 90%-110% of performance as the criteria (Step 3). The reason for taking the prescribed steps to create two control groups is to try to match most of the firms without compromising on the tightness of the matches on performance (Hendricks & Singhal, 2008).

Measurements of Indicators

We measure the fashion and textiles firms' supply chain efficiency by measuring the inventory days, accounts receivable days, and operating cycle time. For the calculation of inventory days, we first divide the cost of goods sold by average inventory. We then divide the 365 days by inventory turnover ratio for the Inventory days. The unit of inventory days is in terms of day (see Formula 1). For accounts receivable days, similarly, we first calculate the accounts receivable turnover by dividing the credit sales by average accounts receivable. We then use 365 days over accounts receivable turnover to estimate the number of accounts receivable days (see Formula 2). The overall operating cycle is the summation of inventory days and accounts receivable days, and it represents the time required to turn the raw materials to cash from customers. The corresponding formulas as follows:

Letting IT be inventory turnover ratio, given by

$$IT = \frac{COGS}{Avg.Inv.},$$

Table 1. Abnormal operating cycles of sample firms for the ISO 9000 adoption in three-year period (year -2 to year 1), four-year period (year -2 to year 2), and five-year period (year -2 to year 3), based on performance-industry-matched and performance-industry-size-matched matching. $p < 0.1$; $p < 0.05$**; $p < 0.01$*** for one-tailed tests; [a] in percentage*

Performance-industry-matched																					
	year - 2 to year 1							year -2 to year 2							year -2 to year 3						
Performance measures	**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**	
Operating Cycle	149	-18.44		-13.24		63.00		142	-11.59		-14.13		63.00		139	-17.24		-15.12		67.00	
t / z statistics		-2.19	**	-3.46	***	-3.11	***		-2.51	***	-3.64	***	-3.11	***		-3.34	***	-4.15	***	-3.90	***
Inventory Days	149	-14.00		-10.31		62.00		142	-11.00		-9.69		63.00		139	-14.48		-9.47		65.00	
t / z statistics		-2.28	**	-3.05	***	-2.79	***		-2.83	***	-3.26	***	-3.11	***		-3.10	***	-3.46	***	-3.39	***
Accounts Re-ceivable Days	149	-9.95		-1.05		51.00		142	-1.78		-1.56		54.00		139	-7.56		-4.32		61.00	
t / z statistics		-1.86	**	-1.71	**	-0.16			-0.80		-1.40	*	-0.92			-2.48	***	-2.76	***	-2.55	***
Performance-industry-size-matched																					
	year - 2 to year 1							year -2 to year 2							year -2 to year 3						
Performance measures	**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**	
Operating Cycle	144	-12.43		-5.51		57.00		137	-9.07		-3.92		59.00		131	-14.22		-7.18		60.00	
t / z statistics		-1.44	*	-1.68	*	-1.58	*		-1.74	**	-1.51	*	-2.05	**		-1.65	*	-1.44	*	-2.10	**
Inventory Days	144	-9.57		-1.89		55.00		137	-11.28		-3.50		55.00		131	-12.54		-5.46		59.00	
t / z statistics		-1.58	*	-1.57	*	-1.08			-2.39	***	-1.68	**	-1.20			-1.86	**	-1.52	*	-1.92	**
Accounts Re-ceivable Days	144	-2.86		0.20		50.00		137	2.32		0.61		47.00		131	-2.41		0.38		48.00	
t / z statistics		-0.88		-0.53		0.00			1.06		1.69		0.68			-0.77		-0.20		-0.35	

Where

COGS = cost of goods sold,
Avg.Inv. = average inventory balance,

We have

$$I = \frac{365}{IT}. \qquad (1)$$

Similarly, letting *ART* be the accounts receivable turnover ratio, given by

$$ART = \frac{CS}{Avg.AR},$$

Where

CS = credit sales,
Avg. AR = average accounts receivable balance,

We have

$$AR = \frac{365}{ART}. \qquad (2)$$

$$OC = I + AR \qquad (3)$$

Where

OC = operating cycle,
I = number of inventory days,
AR = number of accounts receivable days.

We estimate abnormal supply chain efficiency (i.e., inventory days, accounts receivable days, and operating cycle time) within the event window as the difference between sample post-event performance (i.e., actual performance in year 1, year 2 and year 3) and expected performance (in year 1, year 2 and year 3). We estimate expected performance as the sum of sample pre-event performance (i.e., in year -2) and change in the median performance of control firms in that period (i.e., from year -2 to year 1, year 2 and year 3). The formulas are as follows:

$AP_{(t+j)} = PS_{(t+j)} - EP_{(t+j)}$,

$EP_{(t+j)} = PS_{(t+i)} + (PC_{(t+j)} - PC_{(t+i)})$,

Where

AP – abnormal performance,
EP – expected performance,
PS – performance of sample firms,
PC – median performance of control firms,
t – year of ISO 9000 / ISO 14000 certification,
i – starting year of comparison (i = -2),
j – ending year of comparison (j = 1, 2, or 3).

We obtain the performance data from the COMPUSTAT database. Since the first ISO 9000 (ISO 14000) certification was awarded in 1990 (1996) and we need performance data at least two years before certification (year -2) and three years after certification (year 3) for analysis, we obtain performance data covering the period 1988 to 2008.

We conduct the Wilcoxon signed-rank (WSR) test to examine the median abnormal performance. We also carry out the Sign test to determine if the percentage of positive abnormal performance is significantly higher (i.e., higher than 50%). To check for consistency, we further conduct the parametric *t*-test on the mean abnormal performance to ensure that our findings are robust. Table 1 and Table 2 present the results of ISO 9000 and ISO 14000 on supply chain efficiency respectively.

RESULTS

We begin our discussion by examining abnormal supply chain efficiency on ISO 9000 certified textiles and textiles related firms. The cumula-

Table 2. Abnormal operating cycles of sample firms for the ISO 14000 adoption in three-year period (year -2 to year 1), four-year period (year -2 to year 2), and five-year period (year -2 to year 3), based on performance-industry-matched and performance-industry-size-matched matching. $p < 0.1$; $p < 0.05$** for one-tailed tests; [a] in percentage*

Performance-industry-matched																					
	year - 2 to year 1							year -2 to year 2							year -2 to year 3						
Performance measures	**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**	
Operating Cycle	45	-13.36		-11.15		64.00		41	0.30		-3.93		59.00		33	-5.49		-4.37		55.00	
t / z statistics		-1.79	**	-2.13	**	-1.79	**		0.04		-0.93		-0.94			-0.82		-0.87		-0.35	
Inventory Days	45	-6.36		-4.86		64.00		41	-3.76		-6.96		61.00		33	-6.22		-4.17		58.00	
t / z statistics		-1.45	*	-1.93	**	-1.79	**		-0.63		-1.29		-1.25			-1.10		-1.39	*	-0.70	
Accounts Receivable Days	45	-15.45		-2.54		62.00		41	-2.84		-0.14		51.00		33	-3.90		-2.33		58.00	
t / z statistics		-2.05	**	-1.95	**	-1.49	*		-0.54		-0.16		0.00			-0.95		-0.78		-0.70	
Performance-industry-size-matched																					
	year -2 to year 1							year -2 to year 2							year -2 to year 3						
Performance measures	**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**		**N**	**Mean**		**Median**		**% negative**	
Operating Cycle	40	-3.21		-2.99		55.00		37	2.71		-4.66		59.00		32	-4.66		-3.03		53.00	
t / z statistics		-0.68		-0.91		-0.47			0.37		-0.46		-0.99			-0.68		-0.64		-0.18	
Inventory Days	40	-4.78		-2.01		55.00		37	-2.26		-2.64		59.00		32	-5.08		-3.51		56.00	
t / z statistics		-1.15		-1.25		-0.47			-0.33		-1.12		-0.99			-0.89		-1.18		-0.53	
Accounts Receivable Days	40	1.63		3.74		40.00		37	5.03		3.99		43.00		32	-1.01		-1.92		56.00	
t / z statistics		0.45		-0.90		1.11			1.87		1.49		0.66			-0.34		-0.51		-0.53	

tive results of three-year to five-year (from year -2 to year 1, year 2 and year 3) changes provide a clearer picture on the long term impact of ISO 9000 adoption on firms' supply chain efficiency in the fashion supply chain. For the results of the *performance-industry-matched* group from the year prior to ISO 9000 implementation (year -2) to the year of post-certification (year 1), the median (mean) cumulative changes in operating cycle is -13.24 days (-18.44 days), which is significant at the 1% (1%) level. More than half of the sample firms (63%) experience a shorter operating cycle, which is significantly higher than 50% ($p < 0.01$). The median (mean) abnormal change in inventory days is -10.31 days (-14.00 days), which is significant at the 1% (1%) level, with over 62% of the sample firms shortened their inventory days. The median (mean) abnormal change in accounts receivable days is -1.05 days (-9.95 days), which is significant at the 5% (5%) level. More than half of the sample firms (51%) experiencing a shorter accounts receivable days.

For the results of *performance-industry-matched* group from year -2 to year 2, the median (mean) cumulative changes in operating cycle is -14.13 days (-11.59 days), which is significant at the 1% (1%) level. About 63% of sample firms experience a shorter operating cycle, which is significantly higher than 50% ($p < 0.01$). The median (mean) abnormal change in inventory days is -9.69 days (-11.00 days), which is significant at the 1% (1%) level, with over 63% of the sample firms shortened their inventory days. The median (mean) abnormal change in accounts receivable days is -1.56 days (-1.78 days). More than half of the sample firms (54%) experiencing a shorter accounts receivable days.

For the results of *performance-industry-matched* group from year -2 to year 3, the median (mean) cumulative changes in operating cycle is -15.12 days (-17.24 days), which is significant at the 1% (1%) level. About 67% of sample firms experience a shorter operating cycle, which is significantly higher than 50% ($p < 0.01$). The median (mean) abnormal change in inventory days is -9.47 days (-14.48 days), which is significant at the 1% (1%) level, with over 65% of the sample firms shortened their inventory days. The median (mean) abnormal change in accounts receivable days is -4.32 days (-7.56 days). More than half of the sample firms (61%) experiencing a shorter accounts receivable days.

The results of abnormal supply chain performance are similar between the *performance-industry-matched* group and the *performance-industry-size-matched* group, except for the accounts receivable days results. In both matching groups, the impacts of ISO 9000 on operating cycle and inventory days are statistically significant. Hypotheses H1a and H3a are supported. However, the abnormal impact of ISO 9000 on accounts receivable days is not statistically significant in the *performance-industry-size-matched* group. This mixed results suggest that accounts receivable days is sensitive to firm size; hypothesis H2a is only partially supported.

The overall results are robust between the three-year, four-year and five-year cumulative results, revealing that the impact of ISO 9000 on supply chain performance is long lasting in fashion and textiles industries. The length of shorten operating cycle is longer in the period from year -2 to year +3, which means the impact of ISO 9000 is long lasting. The certified textiles related firms improve their supply chain efficiency continuously in the five-year period.

Table 2 presents the results of ISO 14000 matching groups. The *performance-industry-matched* group from the year -2 to year 1, the median (mean) cumulative changes in operating cycle is -11.15 days (-13.36 days), which is significant at the 5% (5%) level. More than half of the sample firms (64%) experience a shorter operating cycle, which is significantly higher than 50% ($p < 0.05$). The median (mean) abnormal change in inventory days is -4.86 days (-6.36 days), which is significant at the 5% (5%) level, with over 64% of the sample firms shortened their inventory days.

The median (mean) abnormal change in accounts receivable days is -2.54 days (-15.45 days), which is significant at the 5% (5%) level. More than half of the sample firms (62%) experiencing a shorter accounts receivable days.

The abnormal median changes of all three supply chain performance indicators are negative in the four-year (from year -2 to year 2) and five-year periods (from year -2 to year 3). However, they are not statistically significant in nearly all the statistical tests. Such results suggest that the impact of ISO 14000 on fashion and textiles related firms' supply chain efficiency is temporary. This impact diminished after year 1. We could not found significant impact of ISO 14000 on the three indicators in the *performance-industry-size-matched* group. These mixed results suggest that hypotheses 1b, 2b and 3b are not supported. The adoption of ISO 14000 only has some short-term impact on supply chain efficiency but it is not enduring. We will discuss these findings in the discussion section.

MODERATING FACTORS OF ISO 9000 ADOPTION IN THE FASHION SUPPLY CHAIN

We try to provide a deeper understanding on the association between ISO 9000 adoption and the improvement in operating cycle. We only focus on ISO 9000 because we reveal no significant change on operating cycle from ISO 14000 adoption in the previous section. We construct a regression model to study how firm-level characteristics affect the impact of ISO 9000 on abnormal operating cycle over the five-year event period (from year -2 to year 3). We use firm size, and original financial performance of the ISO 9000 certified firms as the control factors. The abnormal performance in operating cycle of ISO 9000 certified firms could be more positive for larger firms. Larger firms normally have more resources for hiring external consultants, providing additional training, and dedicating additional manpower to facilitate the implementation process of ISO 9000. We use firms' total assets to represent the firm size.

We also control the financial performance of the firm because firms that are more profitable are more efficient in operations. As ISO 9000 adoption calls for improvement in a firm's operational efficiency, firms that are more efficient may be able to implement the system more effectively than less efficient firms. Firms that are more profitable could also have more resources for ISO 9000 implementation. We estimate the financial performance of a firm as the firm's ROA in year 3.

We include three independent variables into the regression model, which are labour intensity, R&D intensity, and time of ISO 9000 adoption. The arguments and predictions of the three indicators are as follows:

Independent variables:

1. *Labour intensity:* The abnormal changes in operating cycle could be more positive (shortened length of operating cycle) in more labour intensive firms. It is because these firms might have higher need in standardizing their operation procedure to ensure the production processes is smooth. The calculation of labour intensity is the number of employee over firm's total assets.
2. *R&D intensity*: Industries with higher R&D intensity would normally mean that they face more rapid technological and product changes. There are thus more opportunities for new product development for high-technology textiles firms. This allows them to implement efficient process designs to a greater extent. Therefore, the positive impact of ISO 9000 on abnormal operating cycle could be higher in firms with higher levels of R&D intensity. We measure R&D intensity as R&D expenses over sales.
3. *Time of ISO 9000 adoption*: According to the institutional theory, early adoptions of organizational innovations are motivated by

technical and economic needs (DiMaggio & Powell, 1983), while later adopters respond to the growing social legitimacy of the innovations as taken-for-granted organizational structure improvements (Westphal, Gulati, & Shortell, 1997). ISO 9000 is a well-recognized example of institutionalized management practice (Guler, et al., 2002). Therefore, we predict that early adopter of ISO 9000 could have larger improvement in operating cycle than the later ones, as the formers are motivated by technical benefits of ISO 9000.

Table 3 presents the cross-sectional regression results. We use the abnormal performance in operating cycle of the *performance-industry-matched* group for analysis. For this model, the *F*-value is 2.89, which is significant at the 1% level. The adjusted R^2 values is 8.0%, which is comparable to those observed in previous studies that attempted to use cross-sectional regression models to explain abnormal performance (e.g., Hendricks & Singhal, 2008).

We find that firm size and firm labour intensity do not moderate the association between ISO 9000 adoption and abnormal operating cycle. Fashion and textiles firms' ROA is negatively related ($p < 0.01$) to abnormal operating cycle. It means textiles firms that are more profitable can further shortened their operating cycle time from ISO 9000 adoption. The coefficient of R&D intensity is negatively and statistically significant at the 1% level ($p < 0.01$). It means that high technology fashion and textiles firms could benefit more from the reduction of operating cycle time after ISO 9000 adoption. The coefficient of time of adoption is positive and statistically significant at the 5% level. Late adopters of ISO 9000 obtain less abnormal benefit from ISO 9000 adoption (i.e. they have longer abnormal operating cycle compared to early adopters). It shows that the institutionalization of ISO 9000 in general manufacturing industries also appear in the fashion supply chain.

DISCUSSION AND SUMMARY

This study provides empirical evidence on the impact of QMS and EMS adoption on firms' supply chain performance in the fashion supply chain. Based on the 248 ISO 9000 and 61 14000 certified publicly-listed fashion or textiles related

Table 3. Estimated coefficients (t-statistics in parentheses) from regression of abnormal operating cycle change from year -2 to year 3

Independent Variables			
Intercept	-5879	(-1.684)	**
Firm size +	0.001	(0.663)	
ROA +	-108.259	(-2.911)	***
Labour intensity	200.802	(0.939)	
R&D intensity +	-61.501	(-2.463)	***
Year of ISO 9000 adoption +	2.940	(1.683)	**
Number of observations		109	
Model F value		2.89	***
R squared (%)		12.2	
Adjusted R squared (%)		8.0	
Note: Significance levels (One-tailed tests) of independent variables: p<0.05<**;p<0.01***			

firms in the U.S., we find that ISO 9000 adoption has a statistically significant positive impact on operating cycle, inventory days and accounts receivable days. Depending on the methods used to estimate abnormal performance, the median abnormal improvement in operating cycle ranges from 7.18 days to 15.12 days shorter than non-certified firms over a five-year period. In other words, the certified firms save up to 15.12 days to turn raw materials into cash payment from their customers. This suggests that the certified fashions or textiles firms become a more efficient node in the fashion supply chain.

The median improvement in inventory days ranges from 5.46 days to 7.18 days over a five-year period. The results show that the improvement in inventory days is the major reason of the overall improvement in operating cycle. QMS certified firms in the fashion supply chain are more efficient in their production process and inventory management, resulting a shorter inventory days compared to non-certified competitors.

The median improvement in accounts receivable days of *performance-industry-matched* group is 4.32 days. The magnitudes of abnormal changes in accounts receivable days are generally smaller than the changes in inventory days. There are also mixed results of abnormal changes in accounts receivable days between the two matching groups, which mean the impact of ISO 9000 is less significant on this indicator. The possible reason is that accounts receivable days is sensitive to firm size, therefore, the impact of ISO 9000 on it was taken away after firm size had been controlled.

In the cross-sectional analysis, we find that high-tech firms could achieve better supply chain performance than low-tech firms in the fashion supply chain. These high-tech firms in the fashion supply chain are the firms with advanced manufacturing techniques and technology. The results show that the adoption of ISO 9000 could help these firms to further improve their utilization of technological resources and thus improve their supply chain operations. The late adopters of ISO 9000 in the fashion supply chain obtain much less supply chain benefits than the early adopters do. The coefficient is about 2.9, which means the abnormal operating cycle is 2.9 days longer for every one-year delay in adoption. This result supports the institutional explanation of ISO 9000 adoption in the operation management literature, suggesting that the late adopters are not driven by the technical efficiency of a management practice. Another possible explanation is that there are less inefficient non-certified firms in the industry in the later years, therefore the comparative improvement in supply chain efficiency become much less significant.

Why EMS has Just Limited Impact on Firms' Supply Chain Efficiency?

We also investigated that the abnormal changes in operating cycle, inventory days, and accounts receivable days in ISO 14000 certified firms in the fashion supply chain. However, such change is not long lasting (only in year -2 to year 1 period) and only appears in the *performance-industry-matched* group. One possible reason is that some ISO 14000 certified firms sample have already ISO 9000 certified prior the adoption of ISO 14000. Therefore, the incremental benefits from ISO 14000 are smaller than their first ISO 9000 certification. It is also possible that there are only limited number of ISO 14000 firms in our sample to reveal any statistical significance of ISO 14000 impact. In addition, a previous study showed that ISO 9000 is partly driven by bilateral trade relationships in the supply chain, while ISO 14000 is by culture (Albuquerque, et al., 2007), it could be the reason why ISO 9000 could show more benefits in supply chain efficiency comparatively. Although both certifications are being used as supplier selection tools, only ISO 9000 shows strong and significant impact on firms' supply chain performance. The results suggested that the adoption of QMS is likely to be the primary "glue" that makes the nodes in the fashion supply chain works more closely with each other and thus the whole supply chain could become more efficient.

Managerial Implications to Fashion and Textiles Manufacturers

Fashion and textiles firms, which have not obtained ISO 9000, should consider such adoption. It is obvious that ISO 9000 adoption is necessary for improving firms' supply chain efficiency. Our cross-sectional analysis results further suggest that labour intensity is not a moderating factor of ISO 9000 impact on supply chain efficiency. It means that it works not only labour intensive textiles firms (which are very common in textiles industry), but also highly automated textiles firms. For textiles firms with higher R&D intensity, which often have faster new product development and shorter product life cycles, could benefit more from being certified by ISO 9000. In other words, firms that have already adopted ISO 9000 can further improve their supply chain efficiency, by implementing a corporate strategy with higher R&D investment for new products. The late adopters of ISO 9000 should also pay attention to the motivation in obtaining this certification. If the motivation is largely institutional (i.e. by responding to customer requirement), instead of improving firms' operational efficiency, such firms received much less supply chain benefits.

Although we could not reveal a consistent and long-term impact from the ISO 14000 adoption on supply chain performance in this study, we could not rule out the possibility that there are other potential benefits for being certified by ISO 14000. Given the large amount of pollutant emissions and high level of energy and water consumptions during textiles manufacturing processes, the benefits of ISO 14000 might be more on firms' cost efficiency. Therefore, the results should not be interpreted as the lack of usefulness of ISO 14000. Besides, the impact of ISO 14000 on firms' supply chain efficiency on other manufacturing industries might be exists. The mixed results of ISO 14000 in this study should not be implied that ISO 14000 has no impact to general manufacturing industries.

LIMITATIONS AND FUTURE RESEARCH

This study is the first exploratory empirical work focuses on the QMS and EMS adoption in the fashion supply chain and its impact of supply chain performance. There are some limitations should be addressed in future research. First, this study has only included the fashion and textiles related firms in manufacturing industries. However, fashion retailers are also key players in the fashion supply chain. They might have dominating negotiation power over their suppliers in the supply chain. The power between the buyer-supplier relationships may have strong moderating effects on the abnormal performance. Second, we only included all the publicly listed textiles related firms in the fashion supply chain of the U.S. However, most fashion supply chains involve significant number of suppliers from developing countries (e.g., China and India). The impact of ISO 9000 on the textiles firms from developing countries might not be the same as the firms in the U.S. The textiles firms in developing countries might adopt the standard not because of its technical efficiency, but for the legitimacy from its customers in developed countries. Therefore, a replication of the current study under developing countries context could provide the complete picture of the impact of QMS and EMS in the fashion supply chain. Finally, this study only explored the impact of QMS and EMS on supply chain performance. Further studies should be made with the focus on firms' profitability and cost efficiency, to provide additional evidence of the usefulness of these certifications in the fashion supply chain.

REFERENCES

Adanur, S., & Allen, B. (1995). First results on the effects of ISO 9000 in the US textile industry. *Benchmarking for Quality Management & Technology*, *2*(3), 41–52.

Albuquerque, P., Bronnenberg, B. J., & Corbett, C. J. (2007). A spatiotemporal analysis of the global diffusion of ISO 9000 and ISO 14000 certification. *Management Science*, *53*(3), 451–468. doi:10.1287/mnsc.1060.0633

Allen, N., & Oakland, J. S. (1988). Quality assurance in the textile industry: Part I. *International Journal of Quality & Reliability Management*, *5*(5), 25–37. doi:10.1108/eb002916

Allen, N., & Oakland, J. S. (1991a). Quality assurance in the textile industry: Part II. *International Journal of Quality & Reliability Management*, *8*(1), 22–30. doi:10.1108/02656719110141006

Allen, N., & Oakland, J. S. (1991b). Quality assurance in the textile industry: Part III. *International Journal of Quality & Reliability Management*, *8*(2), 9–20. doi:10.1108/02656719110140971

Bansal, P., & Hunter, T. (2003). Strategic explanations for the early adoption of ISO 14001. *The Journal of Business Strategy*, *46*(3), 289–299.

Bansal, P., & Roth, K. (2000). Why companies go green: A model of ecological responsiveness. *Academy of Management Journal*, *43*(3), 717–736. doi:10.2307/1556363

Barber, B. M., & Lyon, J. D. (1996). Detecting abnormal operating performance: The empirical power and specification of test statistics. *Journal of Financial Economics*, *41*(3), 359–399. doi:10.1016/0304-405X(96)84701-5

Boiral, O. (2003). ISO 9000: Outside the iron cage. *Organization Science*, *14*(6), 720–737. doi:10.1287/orsc.14.6.720.24873

Boiral, O., & Sala, J.-M. (1998). Environmental management: Should industry adopt ISO 14001? *Business Horizons*, *41*(1), 57–64. doi:10.1016/S0007-6813(98)90065-9

Brio, J. A. d., Fernandez, E., Junquera, B., & Vazquez, C. J. (2001). Motivations for adopting the ISO 14001 standard: A study of Spanish industrial companies. *Environmental Quality Management*, *10*(4), 13–28. doi:10.1002/tqem.1101

Brito, M. P. d., Carbone, V., & Blanquart, C. M. (2008). Towards a sustainable fashion retail supply chain in Europe: Organization and performance. *International Journal of Production Economics*, *114*(2), 534–553. doi:10.1016/j.ijpe.2007.06.012

Buttle, F. (1997). ISO 9000: Marketing motivations and benefits. *International Journal of Quality & Reliability Management*, *14*(9), 936–947. doi:10.1108/02656719710186867

Christmann, P., & Taylor, G. (2001). Globalization and the environment: Determinants of firm self-regulation in China. *Journal of International Business Studies*, *32*(3), 439–458. doi:10.1057/palgrave.jibs.8490976

Corbett, C. J., & Kirsch, D. A. (2001). International diffusion of ISO 14000 certification. *Production and Operations Management*, *10*(3), 327–342. doi:10.1111/j.1937-5956.2001.tb00378.x

Corbett, C. J., Montes-Sancho, M. J., & Kirsch, D. A. (2005). The financial impact of ISO 9000 certification: An empirical analysis. *Management Science*, *51*(7), 1046–1059. doi:10.1287/mnsc.1040.0358

Delmas, M. A. (2002). The diffusion of environmental management standards in Europe and in the United States: An institutional perspective. *Policy Sciences*, *35*(1), 91–119. doi:10.1023/A:1016108804453

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. doi:10.2307/2095101

Eskew, R. K., & Jensen, D. L. (1996). *Financial accounting* (1st ed.). New York, NY: McGraw-Hill.

Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance*, *50*(1), 131–155. doi:10.2307/2329241

Fatima, M., & Ahmed, E. (2010). Quality management in Pakistan's bedwear industry. *Quality Engineering*, *18*(4), 443–451. doi:10.1080/08982110600860268

Flynn, B. B., Schroeder, R. G., & Sakakibara, S. (1994). A framework for quality management research and an associated measurement instrument. *Journal of Operations Management*, *11*(4), 339–366. doi:10.1016/S0272-6963(97)90004-8

Fresner, J. (1998). Starting continuous improvement with a cleaner production assessment in an Austrian textile mill. *Journal of Cleaner Production*, 6.

Guler, I., Guillen, M. F., & Macpherson, J. M. (2002). Global competition, institutions and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates. *Administrative Science Quarterly*, *47*(2), 207–232. doi:10.2307/3094804

Hendricks, K. B., & Singhal, V. R. (2008). The effect of product introduction delays on operating performance. *Management Science*, *54*(5), 878–892. doi:10.1287/mnsc.1070.0805

ISO. (2004). *ISO 9000 quality management principles*.

ISO. (2009). *The ISO survey - 2008*.

Jennings, P. D., & Zandbergen, P. A. (1995). Ecological sustainable organizations: An institutional approach.

Jiang, R. J., & Bansal, P. (2003). Seeing the need for ISO 14001. *Journal of Management Studies*, *40*(4), 1047–1067. doi:10.1111/1467-6486.00370

Klassen, R. D., & McLaughlin, C. P. (1996). The impact of environmental management on firm performance. *Management Science*, *42*(8), 1199–1214. doi:10.1287/mnsc.42.8.1199

Kojima, M., Nakashima, K., & Ohno, K. (2008). Performance evaluation of SCM in JIT environment. *International Journal of Production Economics*, *115*(2), 439. doi:10.1016/j.ijpe.2007.11.017

Kollman, K., & Prakash, A. (2001). Green by choice? Cross-national variations in firms' responses to EMS-based environmental regimes. *World Politics*, *53*(April), 399–430. doi:10.1353/wp.2001.0010

Kollman, K., & Prakash, A. (2002). EMS-based environmental regimes as club goods: Examining variations in firm-level adoption of ISO 14001 and EMAS in U.K., U.S. and Germany. *Policy Sciences*, *35*(1), 43–67. doi:10.1023/A:1016071810725

Lo, C. K. Y., Yeung, A. C. L., & Cheng, T. C. E. (2009). ISO 9000 and supply chain efficiency: Empirical evidence on inventory and account receivable days. *International Journal of Production Economics*, *118*(2), 367–374. doi:10.1016/j.ijpe.2008.11.010

McGahan, A. M., & Porter, M. E. (1997). How much does industry matter, really? *Strategic Management Journal, 18*(Special Issue: Organizational and Competitive Interactions), 15-30.

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of Management Journal*, *31*(4), 854–872. doi:10.2307/256342

Motwani, J., Youssef, M., Kathawala, Y., & Futch, E. (1999). Supplier selection in developing countries: A model development. *Integrated Manufacturing Systems*, *10*(3/4), 154–161. doi:10.1108/09576069910264411

Naveh, E., & Marcus, A. (2005). Achieving competitive advantage through implementing a replicable management standard: Installing and using ISO 9000. *Journal of Operations Management*, *24*(1), 1–26. doi:10.1016/j.jom.2005.01.004

Pan, J.-N. (2003). A comparative study on motivation for and experience with ISO 9000 and ISO 14000 certification among Far Eastern countries. *Industrial Management + Data Systems, 103*(8/9), 564-578.

Poksinska, B., Dahlgaard, J. J., & Eklund, J. A. E. (2003). Implementing ISO 14000 in Sweden: Motives, benefits and comparisons with ISO 9000. *International Journal of Quality & Reliability Management*, *20*(5), 585–606. doi:10.1108/02656710310476543

Potoski, M., & Prakash, A. (2005). Green clubs and voluntary governance: ISO 14001 and firms' regulatory compliance. *American Journal of Political Science*, *49*(2), 235–248. doi:10.1111/j.0092-5853.2005.00120.x

Quazi, H. A., Khoo, Y.-K., Tan, C.-M., & Wong, P.-S. (2001). Motivation for ISO 14000 certification: Development of a predictive model. *Omega*, *29*(6), 525–542. doi:10.1016/S0305-0483(01)00042-1

Ren, X. (2000). Development of environmental performance indicators for textile process and product. *Journal of Cleaner Production*, *8*(6), 473–481. doi:10.1016/S0959-6526(00)00017-2

Romano, P., & Vinelli, A. (2001). Quality management in a supply chain perspective. Strategic and operative choices in a textile-apparel network. *International Journal of Operations & Production Management*, *21*(4), 446–460. doi:10.1108/01443570110381363

Russo, M. V., & Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal*, *40*(3), 534–559. doi:10.2307/257052

Sarkar, A. (1998). Implementation of ISO 9000 in a textile mill. *total Quality Management & Business Excellence, 9*(1), 123-131.

Stewart, G. (1995). Supply chain performance benchmarking study reveals keys to supply chain excellence. *Logistics Information Management*, *8*(2), 38–44. doi:10.1108/09576059510085000

Stewart, G. (1997). Supply-chain operations reference model (SCOR): The first cross-industry framework for integrated supply-chain management. *Logistics Information Management*, *10*(2), 62–67. doi:10.1108/09576059710815716

Teng, S. G., & Jaramillo, H. (2005). A model for evaluation and selection of suppliers in global textile and apparel supply chains. *International Journal of Physical Distribution & Logistics Management*, *35*(7/8), 503–521. doi:10.1108/09600030510615824

Thaver, I., & Wilcock, A. (2006). Identification of overseas vendor selection criteria used by Canadian apparel buyers. Is ISO 9000 relevant? *Journal of Fashion Marketing and Management*, *10*(1), 2006. doi:10.1108/13612020610651123

Westphal, J. D., Gulati, R., & Shortell, S. M. (1997). Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. *Administrative Science Quarterly*, *42*(2), 366–394. doi:10.2307/2393924

KEY TERMS AND DEFINITIONS

Accounts Receivable Days: It is an average number of days a firms needed to receive payments from their customers.

Accounts Receivable Turnover: It is the ratio of the number of times that accounts receivable is collected in one year.

Inventory Days: It is the average number of days goods remain in inventory before being sold to customers.

Inventory Turnover: It is the ratio showing how many times a firm's inventory is sold and replaced in one year.

ISO 14000: An environmental management system developed by ISO in 1996, and its latest version is ISO 14000:2004

ISO 9000: A quality management system developed by ISO in 1987, and its latest version is ISO 9000:2008.

Operating Cycle: It is the time required from turning raw materials into products to customers pay for their accepted products, it is also called cash-to-cash cycle.

This work was previously published in Fashion Supply Chain Management: Industry and Business Analysis, edited by Tsan-Ming Choi, pp. 111-129, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 4

People-Focused Knowledge Sharing Initiatives in Medium-High and High Technology Companies:

Organizational Facilitating Conditions and Impact on Innovation and Business Competitiveness

Nekane Aramburu
University of Deusto, Spain

Josune Sáenz
University of Deusto, Spain

ABSTRACT

The aim of this chapter is to analyse the organizational conditions that foster the development of different people-focused knowledge sharing initiatives in medium-high and high technology companies, as well as the degree of influence of those initiatives on the ideation stage of innovation processes. Finally, considering that successful innovation is the one that helps to improve business competitiveness, the degree of influence of this innovation capability dimension on company performance is examined. For these relationships to be tested, an empirical study has been carried out among medium-high and high technology Spanish manufacturing firms with more than 50 employees and which carry out R&D activities. To this end, a questionnaire has been designed and submitted to the CEOs of the companies making up the target population of the research. Structural equation modelling (SEM) based on partial least squares (PLS) has then been applied in order to test the main hypotheses of the research.

DOI: 10.4018/978-1-4666-1945-6.ch004

INTRODUCTION

Since the last decade, the study of knowledge has been one of the most important topics in the management arena (Nonaka, 1991; Nonaka & Takeuchi, 1995; Davenport & Prusak, 1998). Indeed, we are supposed to live in a "knowledge economy", where intangible assets – and knowledge in particular – are the key sources for value creation (Brooking, 1996; Edvinsson & Malone, 1997; Stewart, 1997).

This idea is clearly reinforced by the existing relationship between knowledge creation and innovation (Nonaka, 1991; Nonaka & Takeuchi, 1995). In today's economy, innovation is one of the main driving forces behind business competitiveness (Drucker, 1988; Shapiro & Varian, 1998; Sveiby, 1997). Along these lines, it is generally assumed that innovation depends on the accumulation and development of relevant knowledge of a wide variety (Fischer, 2001).

For new knowledge to be created, knowledge sharing between individuals is the key (Nonaka, 1991; Nonaka & Takeuchi, 1995; Nonaka, von Krogh & Voelpel, 2006). As a consequence, the study of different mechanisms and initiatives which could facilitate knowledge sharing represents an extremely relevant research topic. Many of these mechanisms take advantage of information and communication technologies (i.e. they are "IT-based" – Dalkir, 2005; Davenport, 2007) whereas, in other cases, personal interaction between individuals is the key (i.e. "people-focused" knowledge management; Wiig, 2004). A review of early literature on knowledge management gives clear proof of the prevalence of information technology (IT) focused research in this domain (Swan, Robertson & Nevell, 2001).

Taking this into consideration, the focus of this chapter will be on people-focused knowledge sharing (i.e. the type of knowledge sharing which involves personal or "face-to-face" interaction). In particular, the organizational conditions that foster the development of different people-focused knowledge sharing initiatives (e.g. communities of practice, coaching, mentoring, employee functional rotation and other initiatives for knowledge sharing with external stakeholders) will be analysed, as well as the degree of influence of those initiatives on the ideation stage of innovation processes. Finally, considering that successful innovation is the one that helps to improve business competitiveness, the degree of influence of this innovation capability dimension on company performance will be examined. In other words, it is assumed that business competitiveness is related to superior performance (Cantwell, 2005).

Given their special focus on innovation, their extremely high knowledge intensity, and the degree of complexity of the knowledge being dealt with, medium-high and high technology companies will be under scrutiny in this research. As a result, these companies will be provided with a basic framework in order to shape their knowledge management strategies and in order to enhance their capability for generating new ideas and developing successful innovation.

THEORETICAL FOUNDATIONS

The Concept and Nature of Knowledge

A single definition of "knowledge" does not exist, but it is quite common to approach this concept by starting out from the hierarchical distinction between *data*, *information* and *knowledge* highlighted by Davenport and Prusak in 1998. According to these authors, *data* is a set of discrete, objective facts about events; *information* is a message, usually in the form of a document or audible or visible communication; and *knowledge* is a fluid mix of framed experiences, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information.

Usually, a distinction is made between *tacit* and *explicit* knowledge. The concept of *tacit knowledge* was first coined by the philosopher Michael Polanyi in 1966, but it is thanks to the seminal works by Nonaka in 1991 and Nonaka and Takeuchi in 1995 that it has become extremely popular in management literature as well. However, the meaning attributed by Nonaka to this concept differs from the one attributed by Polany (Allee, 2003). For the latter, the tacit dimension of knowledge refers to innate intelligence, perception and capacities for reasoning, whereas for Nonaka, *tacit knowledge* is the type of knowledge which is personal, context-specific and, therefore, hard to formalize and communicate. Conversely, *explicit or codified knowledge* refers to knowledge that is transmittable in formal, systematic language (Nonaka & Takeuchi, 1995). Hence, according to this perception, tacit knowledge tends to reside within the head of knowers, whereas explicit knowledge is usually contained within tangible or concrete media (Dalkir, 2005).

In Nonaka's view, tacit and explicit knowledge are not totally separate, but mutually complementary entities: human knowledge is created and expanded through social interaction between tacit and explicit knowledge. This interaction is called "knowledge conversion" and there are four types of it: from tacit to tacit (socialization); from tacit to explicit (externalization); from explicit to explicit (combination); and from explicit to tacit (internalization). Therefore, the possibility exists of transforming one type of knowledge (tacit or explicit) into the other. However, if the concept of tacit knowledge were the one advocated by Polanyi, such a possibility would not exist. As Allee (2003) points out, for Polanyi, tacit knowledge could never be made explicit. For him, when knowledge is shared, there is an articulated or explicit communication and an unspoken tacit communication going on at the same time.

In any case, Nonaka's point of view is the most widespread one in the knowledge management literature today. According to him, the tacit/explicit interaction is continuous and dynamic and is shaped by shifts between the different modes of knowledge conversion, which gives rise to a "knowledge creation spiral". As previously mentioned, *socialization* involves the conversion of tacit knowledge into a tacit one. This only can be achieved by a process of experience sharing. As a result of this, a set of shared mental models and technical skills will be obtained. In *externalization*, tacit knowledge is articulated into explicit concepts, using metaphors, analogies, hypotheses or models. This is triggered by dialogue or collective reflection. On the other hand, *combination* involves systemizing concepts into a knowledge system, which implies using different bodies of explicit knowledge. Documents, meetings, conversations or computerized communication networks could be used to this end. Finally, *internalization* is closely related to the idea of "learning by doing", and it means embodying explicit knowledge into (a) tacit one. "For explicit knowledge to become tacit, it helps if the knowledge is verbalized or diagrammed into documents, manuals or oral stories" (Nonaka & Takeuchi, 1995, p. 69).

As can be seen, in all the previously-mentioned processes knowledge sharing is involved, which means this is a critical aspect in enlarging organizational knowledge. In other words, the knowledge that the organization possesses cannot be amplified if the knowledge possessed by individuals is not shared. This shows us the social nature of knowledge. For this reason, different authors have been interested in the study of knowledge sharing mechanisms in organizations (Nahapiet & Ghoshal, 1998; Nonaka, Toyama & Byosière, 2003). In the next section, we are going to take a closer look at this issue.

Knowledge Sharing

As previously mentioned, in order to make knowledge sharing happen within and among organizations, several *ad hoc* mechanisms and initiatives have been proposed. Some of them are IT-based (email; on-line discussion forums and/or blogs; intranets; extranets; groupware tools; and on-line knowledge repositories), whereas in other cases, personal interaction between individuals is the key (communities of practice, coaching, mentoring, and employee functional rotation, to name but a few).

Early literature on knowledge management shows a clear prevalence of the information technology perspective. In a survey reported by Swan, Robertson & Newell in 2001, it was proven that the dominant discourse in this domain was not related to the role of people in organizations, but to technologies that facilitate the acquisition, codification and exploitation of knowledge. However, although IT-based practices could help in the exchange and assimilation of explicit knowledge (i.e. internalization and combination processes), they leave aside "social construction" (Berger & Luckmann, 1966) and the organizational learning processes underlying socialization. As Allee points out (2003), "in all types of knowledge work, even where technology is very helpful, people require conversation, experimentation and experiences shared with other people who do what they do" (p. 113).

In this chapter, the focus will be on this second perspective: that is, on social interaction-based (i.e. people-focused) knowledge sharing. In particular, the following knowledge sharing initiatives will be considered: communities of practice, coaching and mentoring, employee functional rotation and meeting events and/or workshops in order to promote reflection as well as knowledge and experience sharing with external agents.

Among the previously-mentioned mechanisms, communities of practice are of special relevance (Dalkir, 2005; Saint-Onge & Wallace, 2003). They refer to a group or network of individuals who share a concern, a set of problems or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting with each other on an ongoing basis (Wenger, McDermott & Snyder, 2002). Along these lines, "the critical component of a community lies in the sharing of common work problems between members, a membership that sees the clear benefits of sharing knowledge among themselves" (Dalkir, 2005, p.123).Therefore, communities of practice are a vehicle for increasing knowledge creation as well as for expanding the extent and accelerating the speed at which knowledge is exchanged across the organization (Saint-Onge & Wallace, 2003, p.12).

On the other hand, coaching, mentoring and employee functional rotation facilitate learning and socialization processes across the organization as well. In particular, they are used to allow the transmission of knowledge that is grounded on experience (i.e. tacit knowledge). According to action learning theory (Revans, 1983), individuals learn through experience within a social context. This type of experience-grounded learning is quite difficult to transmit to others because of its tacit nature. For this reason, a social context is needed to allow this transmission. As work-based learning theory maintains (Raelin, 2000), individuals learn in their work environment, which constitutes the immediate social context in which they gain experience. In this context, individuals not only gain experience, but also share it with others. Hence, socialization processes (i.e. tacit knowledge sharing) take place.

Finally, the mobilization of external knowledge held by outside stakeholders is also an essential aspect for knowledge creation to occur (Nonaka & Takeuchi, 1995; Almeida, Anupama & Grant, 2003). In other words, the exchange of knowledge with external agents is a key element for creating new knowledge. As Maznevski & Athanassiou (2007) point out, "the scope and breadth of knowledge available from outside sources is generally much greater than that available from inside sources" (p. 69). Therefore, meeting events and/or workshops in order to promote reflection as well as knowledge and experience sharing with external agents constitute other relevant social interaction-based mechanisms to enable knowledge sharing and subsequent knowledge creation to take place.

On the other hand, for knowledge sharing to occur, specific spaces should be created within the organization, where knowledge exchange among people would be facilitated. These particular spaces are what Nonaka *et alter* called "ba" (Nonaka, Reinmoeller & Senoo, 1998). According to these authors, the ba is the physical and/or virtual space where knowledge exchange and generation take place. As stated by Nonaka, Toyama & Hirata (2008), it is "an existential place where participants share contexts and create new meanings through interactions"(p.34).

More precisely, the *physical ba* refers to those organizational conditions which can promote or, on the contrary, hinder the exchange of knowledge. In particular, organizational design and technological infrastructure are included within this category (Nonaka & Takeuchi, 1995). On the other hand, *the virtual ba* (i.e. the mental space for knowledge sharing) is linked to organizational culture or the set of shared values assumed by the members of the organization. Sharing a specific set of values generates a collective mental space which facilitates mutual understanding among people and, hence, mutual interaction and knowledge sharing.

In this research, the influence of different organizational conditions (i.e. ba components) on people-based knowledge sharing is analysed.

Knowledge Sharing and its Relationship with Knowledge Creation and Innovation

Since the seminal works by Nonaka in 1991, and Nonaka & Takeuchi in 1995, the concept of innovation has been closely related to that of "knowledge creation". Along these lines, it is generally assumed that the process of innovation consists of an ongoing pursuit of harnessing new and unique knowledge (Subramaniam & Youndt, 2005).

According to Nonaka, von Krogh and Voelpel (2006), knowledge creation involves a continuous process through which one overcomes the individual boundaries and constraints imposed by information and past learning by acquiring a new context, a new view of the world and new knowledge. By interacting and sharing tacit and explicit knowledge with others, the individual enhances the capacity to define a situation or problem, and apply his or her knowledge so as to act and specifically solve the problem. In the case of organizational knowledge creation, this means making available and amplifying the knowledge created by individuals as well as crystallizing and connecting it with the organization's knowledge system (Nonaka & Takeuchi, 1995; Nonaka, von Krogh & Voelpel, 2006). Therefore, knowledge sharing and diffusion are both essential in order to create new knowledge and produce innovation (Dalkir, 2005).

On the other hand, innovation is a dynamic capability (i.e. a capability which allows the company to create, extend or modify its resource base – Helfat *et alter*, 2007) with multiple dimensions. The first one is that of ideation (Davila, Epstein & Shelton, 2006), which involves new idea generation and selection. Selected new ideas should then be put into practice (Van de Ven & Angle, 2000). This brings us to the execution phase. In this phase, managing innovation projects effectively should be combined with the ability to fit them into budgeted costs and deadlines (i.e. timeliness and cost efficiency). Finally, all the aforementioned dimensions should lead to value creation.

Considering that the generation and selection of new ideas is the first step in order to achieve successful innovation and that this is something not specifically addressed by previous research (which in many cases assesses innovation capability in terms of the outcomes it produces, instead of measuring the capability itself), in this chapter the focus will be on the ideation phase of innovation processes and on value creation. In particular, the impact of people-focused knowledge sharing initiatives on the generation of new ideas will be examined, as well as the impact of the latter on firm performance.

DEVELOPING A RESEARCH MODEL FOR ANALYSING THE INFLUENCE OF ORGANIZATIONAL CONDITIONS ON PEOPLE-FOCUSED KNOWLEDGE SHARING AND THE IMPACT OF THE LATTER ON THE GENERATION OF NEW IDEAS AND ON COMPANY PERFORMANCE

In this section, a research model will be proposed in order to analyse the degree of influence of different organizational dimensions (i.e. organizational design, organizational culture and information and communication technology infrastructure) on the degree of success of different people-focused knowledge sharing initiatives and their impact on the generation of new ideas and company performance.

This model will allow us to answer the following questions:

- Are all organizational dimensions considered equally important in order to support people-focused knowledge sharing initiatives?
- What is the degree of influence of those initiatives when it comes to enhancing the generation of new ideas and, therefore, the innovation capability of firms?
- What is the degree of relevance of this dimension (i.e. ideation) in successful innovation (i.e. in the improvement of business competitiveness)?

The answers to these questions will provide companies with important clues so as to shape their knowledge management and innovation strategies.

As regards the different knowledge sharing catalysts considered, the first one is organizational design. This refers to the type of organizational structure in place within the company, to the communication channels (both vertical and horizontal) that link different organizational units and teams, and to the physical design of the workplace.

As regards organizational structure, Nonaka and Takeuchi (1995), and Nonaka, Toyama and Byosière (2003) advocate the fact that certain types of structure facilitate knowledge sharing and knowledge creation processes more than others (i.e. they are more "learning supportive"). In particular, they defend the hypertext type of organization as the most suitable one in order to foster knowledge sharing and creation.

Likewise, communication channels could play a substantial role in fostering knowledge sharing and subsequent knowledge creation. As Kalla (2005) points out, knowledge sharing is a function of integrated internal communications. Hence, it is assumed that vertical and horizontal communication channels act as catalysts for knowledge sharing.

Finally, physical design of the workplace is the last element making up organizational design that could promote or, on the contrary, hinder the knowledge sharing processes. According to Nonaka, Schamer and Toyama (2001), "the single most important factor shaping the quality of knowledge is the quality of place" (p. 233).

In accordance with the prominent role that, from a theoretical point of view, organizational design could play in knowledge sharing, the following hypothesis has been formulated:

H1: Organizational design acts as a catalyst for knowledge sharing that takes place through different people-focused initiatives.

Organizational culture is the second organizational dimension considered that, according to literature, could have a significant influence on knowledge sharing. As Dalkir (2005) points out, "corporate culture is a key component in ensuring that critical knowledge and information flow within an organization" (p. 185). In particular, the creation of an organizational culture based on trust and commitment on the part of individuals is essential (Argyris, 1990; Argyris & Schön, 1978, 1996; Allee, 2003; Friedman, Lipshitz & Overmeer, 2003; Wiig, 2004). All this gives rise to the second research hypothesis:

H2: Organizational culture is a catalyst for knowledge sharing that takes place through different people-focused initiatives.

Finally, information and communication technologies can also contribute to a great extent to knowledge sharing (Davenport, 2007). According to Allee (2003), "there must be a technology infrastructure in place that really supports the right kind of conversations and connections" (p.89). In particular, the existence of specific technological tools that foster the capturing and storing of knowledge, as well as the connection between individuals and groups(,) may be very helpful (Dalkir, 2005). Therefore, the following hypothesis has been formulated:

H3: Information and communication technologies facilitate knowledge sharing that takes place through different people-focused initiatives.

Additionally, and considering the theoretical foundations explained in previous sections, the following hypotheses have been formulated:

H4: People-focused knowledge sharing initiatives have a positive impact on the generation of new ideas.

H5: Managing the generation of new ideas effectively improves business competitiveness.

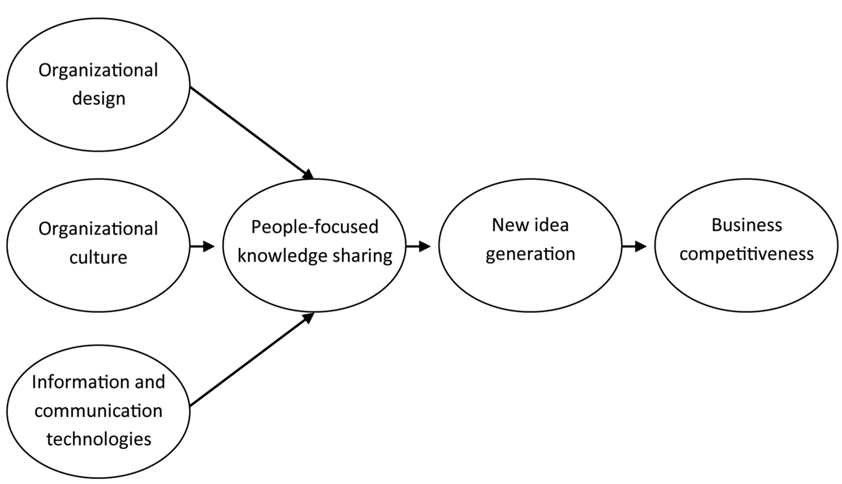
Figure 1 summarizes the model proposed.

TESTING THE RESEARCH MODEL

Research Method

The research model previously outlined has been tested in a sample of Spanish firms. In particular, the population subject to study was made up of medium-high and high technology Spanish manufacturing firms with over 50 employees and which carry out R&D activities. The companies making up the target population were identified thanks to the use of the SABI data base ("Sistema de Análisis de Balances Ibéricos" / System of Iberian Balance Sheet Analysis), which contains the registered annual accounts of over 190,000

Figure 1. Research model



Spanish companies, selecting only those firms which had included their expenses on R&D in their balance sheet.

In order to gather information about the relevant variables of the research, a questionnaire was designed and submitted to the CEOs of the companies making up the target population by the end of the year 2006 and the beginning of the year 2007. 75 answers out 446 were obtained, which means an average response rate of 17%.

The sample size obtained is large enough to carry out a statistical study based on structural equation modelling (partial least squares approach) by means of PLS-Graph software (Chin & Frye, 2003). According to the complexity level of the model to be tested, the minimum sample size required was calculated, and this was made up of 40 firms.

Structural equation modelling (SEM) constitutes a second generation of multivariate analysis which combines multiple regression concerns (by examining dependency relationships) and factor analysis (by representing unobserved variables by means of multiple observed measures), in order to estimate a set of dependency relationships which are all simultaneously interrelated.

When applying SEM, two approaches can be used: the covariance-based approach and the partial least squares (PLS) approach. In the first case, the aim is to minimize the difference between the sample covariances and those predicted by the model. This approach is mainly used for confirmatory analysis. In the second case, however, the aim is to obtain determinate values of the latent variables for predictive purposes. This approach is very useful for exploratory research, as is the case in this chapter (Wold, 1985).

A PLS model is analysed and interpreted in two stages: firstly, the assessment of the reliability and validity of the measurement model and secondly, the assessment of the structural model. This sequence ensures that the constructs' measures are valid and reliable before attempting to draw conclusions regarding relationships among constructs (Barclay *et al.*, 1995).

In the next section, the measurement model is presented in more detail.

Constructs and Measures

The exogenous constructs of the research are those related to the different organizational enablers considered. The first one within this category is *Organizational design*. This construct is made up of four formative indicators which give rise to the existence of an organizational design which favours organizational learning and knowledge sharing. The first indicator refers to the type of organizational structure in place. For the purposes of this research, organizational structures have been classified into three different categories: level 1 learning supportive (i.e. the least learning supportive) organizational structures (that is, functional, divisional or matrix-type structures with no process- or project-based axis); level 2 learning supportive organizational structures (that is, process-based structures or matrix-type structures with a process-based axis); and level 3 learning supportive (i.e. the most learning supportive) organizational structures (that is, project-based structures or matrix-type structures with a project-based axis). The second and third indicators refer to the extent to which vertical and horizontal communication channels allow the flow of ideas, initiatives and points of view in an agile and fluid way, whereas the last one measures the extent to which the physical design of the work environment favours communication and dialogue among all the members of the company. These last three indicators have been measured by means of 1 to 7 Likert scales.

Organizational culture is the second exogenous construct of the research. As has been explained in the theoretical framework, this is linked to the degree of presence within the company of different values and attitudes which are related to a culture of knowledge sharing. The degree of presence of the aforementioned values and attitudes (trust, transparency, open mentality, mistakes considered as learning opportunities, and cooperation and mutual help) has been measured by means of 1 to 7 Likert scales. All the indicators considered are reflective in nature, as they show the type of organizational culture in place within the company.

Finally, *Information and communication technologies* encompass three reflective indicators that show the extent to which the company is equipped with ICT systems which facilitate knowledge sharing and permanent connection with different agents. In particular, these indicators refer to the extent to which ICT systems in place facilitate the storage of organizational knowledge and its easy access by the members of the company; to the extent to which they allow permanent connection among all members of the organization; and to the extent to which they facilitate continued action and joint work with external agents. All of these indicators have been measured by means of 1 to 7 Likert scales.

On the other hand, *People-focused knowledge sharing* is the first endogenous construct of the research. In this case, the degree of use of different mechanisms in order to promote social interaction among individuals has been measured (Wiig, 2004): communities of practice and/or meetings by field of interest; coaching and/or mentoring; employee functional rotation; and meeting events and/or workshops in order to promote reflection as well as knowledge and experience sharing with external agents. All of these indicators have been measured by means of 1 to 7 Likert scales and are formative in nature: that is, the use of the different initiatives mentioned gives rise to the existence of knowledge sharing (i.e. the latent variable being studied).

New idea generation is then the second endogenous construct of the research. In this case, the following items have been checked, all of them referring to the last 5 years: whether the company has been able to identify numerous opportunities for incremental/radical improvement; whether the firm has been able to identify many alternative and new uses for already existing technologies; whether the new idea generation process has been managed in a conscious and effective way; and whether the company has been able to clearly distinguish which of the new opportunities identified had a greater potential for development. Once more, these indicators have been measured by means of 1 to 7 Likert scales and, in this case, they are reflective in nature: that is, they reflect the effectiveness of the new idea generation process.

To bring the presentation of the measurement model to a close, only the *Business competitiveness* construct remains to be explained. With a 5 year scope and using again 1 to 7 Likert scales, the following reflective measures have been chosen: whether incremental/radical innovation projects carried out have shown expected results; whether innovation outcomes have had a very positive impact on the company's income statement; whether innovation outcomes have had a very positive impact on the company's competitive position; and whether innovation outcomes have allowed the company to grow and improve its market share.

MEASUREMENT MODEL EVALUATION

Following the sequence previously described in the method section, an analysis of the results obtained should begin with the evaluation of the measurement model. This assessment differs depending on the nature of the construct under scrutiny. In the case of constructs made up of reflective indicators (i.e. when the measures observed are the consequence of the latent variable and, therefore, should be highly correlated), individual item reliability, construct reliability, convergent validity, and discriminant validity should be checked. However, in the case of constructs made up of formative indicators (i.e. when the measures observed give rise to the existence of the latent variable), multicolinearity problems should be explored.

As regards the reflective constructs of the model (the ones related to organizational culture, information and communication technologies, new idea generation and business competitiveness) all the tests carried out have shown satisfactory results. Indicator loadings (individual item reliability) are greater than 0.7, with a single exception: the one referring to the extent to which the firm has been able to identify many alternative and new uses for already existing technologies, whose loading is 0.6791. Given that the aforementioned value is not too far from the right limit, the decision has been made to keep the indicator in the model.

On the other hand, composite reliability (which measures construct reliability) is higher than 0.8 in all cases; average variance extracted (which measures convergent validity) is greater than 0.5 in all constructs; and discriminant validity is excellent too. As regards formative constructs (i.e. the rest of the constructs of the model), multicolinearity problems have not been identified.

Once the quality of the measurement model has been guaranteed by means of the previously-mentioned tests, the quality of the structural model should then be assessed. This refers to the strength of the research hypotheses and to the amount of variance explained (R^2) in the case of endogenous constructs, as well as to an analysis of the predictive power achieved.

STRUCTURAL MODEL EVALUATION

In order to assess the research hypotheses, path coefficient levels should be examined, as well as their degree of significance, by means of bootstrapping techniques. Tables 1, 2 and 3 summarize the results obtained. In these tables, we can also see the contribution of each exogenous construct to the amount of variance explained (which has been obtained by multiplying correlation and path coefficients), as well as the predictive power achieved. The latter has been confirmed by means of a Stone Geiser test, where cross-validated redundancy (Q^2) must be higher than 0 in order to consider that the model has predictive power for that specific construct.

According to the figures contained in Table 1, we can conclude that the three organizational factors considered (organizational design, orga-

Table 1. Structural model evaluation – Impact of organizational conditions on people-focused knowledge sharing

	Organiza-tional design	Organiza-tional culture	ICT systems	Total amount of variance explained	Predictive power Q^2
Path	0.241*	0.334***	0.248**		
Correlation	0.521	0.544	0.450		
R^2	**12.56%**	**18.17%**	**11.16%**	**41.89%**	**0.0853**

Notes***p<0.001, **p<0.01, *p<0.05 (based on t_{499}, one-tailed test).

Table 2. Structural model evaluation – Impact of people-focused knowledge sharing on the generation of new ideas

Path coefficient	Total amount of variance explained R^2	Predictive power Q^2
0.637***	40.56%	0.1396

Notes***p<0.001 (based on t_{499}, one-tailed test).

Table 3. Structural model evaluation – Impact of the effectiveness of the new idea generation process on business competitiveness

Path coefficient	Total amount of variance explained R^2	Predictive power Q^2
0.630***	39.69%	0.1384

Notes***$p<0.001$ (based on t_{499}, one-tailed test).

nizational culture and information and communication technologies) contribute to a great extent to the knowledge sharing which takes place through different people-focused initiatives in medium-high and high technology firms (amount of variance explained: 42%). Indeed, all path coefficients are statistically significant and, hence, hypotheses 1, 2 and 3 are clearly supported.

As regards the degree of relevance of each explanatory factor in order to enhance this type of knowledge sharing, the most relevant one is organizational culture (contribution to the amount of variance explained: 18%), followed by organizational design (13%) and by information and communication technologies (11%).

On the other hand, people-focused knowledge sharing proves to be extremely important in order to enhance the generation of new ideas in medium-high and high technology firms. As can be seen in Table 2, this type of knowledge sharing is able to explain 41% of the variance in the case of the new idea generation construct and, accordingly, the path coefficient linking both constructs (people-focused knowledge sharing and ideation) is extremely significant. Therefore, hypothesis 4 is clearly supported.

Finally, Table 3 shows that good performance at the ideation stage accounts for 40% of total innovation performance (i.e. business competitiveness improvement). Once more, the path coefficient linking both constructs (ideation and business competitiveness) is statistically significant and, hence, hypothesis 5 is clearly supported.

As a conclusion, people-focused knowledge sharing is of high relevance even in a type of company where technology (and, therefore, IT-based knowledge sharing) is supposed to be at the forefront. This reinforces the relevance of the social dimension of knowledge and the need to facilitate conversation, experimentation and experience sharing among people.

Once the strength of the research hypotheses and the predictive power of the model have been examined, for those explanatory constructs made up of formative indicators (i.e. organizational design and people-focused knowledge sharing), it would be interesting to see the degree of relevance of each specific element making up the construct. For this to be done, indicator weights should be examined. Technically speaking, an analysis of the aforementioned weights should be carried out within the measurement model evaluation section. However, from an intuitive point of view, it seems more logical to know the degree of relevance of the construct as a whole first and, once this has been established, to study the degree of relevance of the specific elements making up the construct in depth. In the case of reflective variables, this analysis does not make any sense, as all the indicators of the construct should be highly correlated and, therefore, their degree of influence should be very similar.

Table 4 shows us the degree of relevance of each specific item making up organizational design, whereas Table 5 does the same thing for people-focused knowledge sharing initiatives.

As can be seen in Table 4, and as far as organizational design is concerned, physical design of the workplace is clearly the most influential item within this construct when it comes to supporting people-focused knowledge sharing. The agility and fluidity of vertical and horizontal com-

Table 4. Organizational design – Ranking of indicators according to their weight

Elements	Weights
Physical design of the workplace	0.5974
Vertical communication channels	0.4610
Horizontal communication channels	0.2922
Organizational structure	0.0081

Table 5. People-focused knowledge sharing – Ranking of indicators according to their weight

Elements	Weights
Communities of practice	0.5786
Employee functional rotation	0.4280
External networking	0.2182
Coaching and/or mentoring	0.1740

munication channels is quite important too, specially in the case of the vertical ones. On the contrary, the type of organizational structure in place is a completely unimportant issue. This is a surprising and unexpected result. In fact, according to theory and previous empirical studies carried out in other geographical settings, organizational structure constitutes a very relevant organizational dimension in order to enhance knowledge sharing. At least this is the case for the Japanese companies studied by Nonaka and Takeuchi in 1995.

As regards people-focused knowledge sharing, communities of practice appear to be the most relevant mechanism in order to enhance the generation of new ideas within medium-high and high technology firms, followed quite closely by employee functional rotation. On the other hand, coaching and/or mentoring as well as the development of different initiatives in order to facilitate knowledge and experience sharing with other stakeholders occupy a more secondary position.

CONCLUSION AND PRACTICAL IMPLICATIONS

The research carried out shows that organizational conditions (i.e. organizational design, organizational culture and information and communication technologies) are extremely important catalysts when it comes to enhancing people-focused knowledge sharing in medium-high and high technology firms. Moreover, this type of knowledge sharing proves to be essential in order to facilitate the generation of new ideas, which is an extremely important dimension of the innovation capability of firms in order to guarantee success in this domain (i.e. in order to improve business competitiveness).

This is a very interesting conclusion of the research. Considering that the companies analysed are technology(-)intensive, it could be thought *a priori* that IT-based knowledge sharing would be the prevalent one or at least that it could "eclipse" social interaction-based knowledge sharing. Nevertheless, the research shows that people-based knowledge sharing is really important in these companies and that this is an essential ingredient for fostering the generation of new ideas.

From a practical point of view, the results obtained provide companies with a useful insight in order to prioritize their knowledge management efforts. In the case of the firms under study (medium-high and high technology firms), developing and nurturing communities of practice as well as supporting employee functional rotation seem to be the best initiatives in order to facilitate the generation of new ideas and the development of innovation. Therefore, communities of practice constitute a very important mechanism when it comes to fostering creativity and to increasing the innovation capability of firms. For this reason, communities of practice should be considered seriously as a key instrument in the implementation process of innovation strategies in medium-high and high technology firms.

On the other hand, and as regards the organizational conditions considered in the research, taking care of the physical design of the workplace and of the agility and fluidity of vertical and horizontal communication channels constitute the two main aspects to be managed within organizational design, together with organizational culture and information and communication technologies.

FUTURE RESEARCH PATHS

The research carried out has shed some light on the influence of different organizational enablers on social interaction-based knowledge sharing in medium-high and high technology companies, and on the influence of the latter on the ideation stage of innovation processes. Moreover, it has shown the relevance of this innovation capability dimension in the final success of innovation processes. Nevertheless, different aspects remain still uncovered, paving the way for future research.

For instance, another organizational factor that could significantly influence knowledge sharing is the assessment and reward system implemented in the organization. In order to foster knowledge sharing it is crucial that this system supports an organizational culture based on trust and commitment on the part of individuals (Allee, 2003). If the reward system is exclusively geared towards individual achievement, it could be difficult to find any incentive to share knowledge with others, as this could slow down personal performance. In the same way, leadership style (participative vs. authoritarian) could also support or, on the contrary, hinder knowledge sharing. Finally, the existence of an explicit knowledge management strategy and the implementation of specific control/assessment tools in this domain could help to a great extent to improve the degree of success of the knowledge sharing initiatives implemented by the company. Therefore, the research model should be enlarged in order to encompass these elements too.

On the other hand, it would be interesting to analyse the influence of this type of knowledge sharing mechanisms in other dimensions of the innovation capability of firms, such as innovation project management.

Additionally, the research carried out could be complemented with the inclusion of different contingent variables, such as company size or industry type. Only manufacturing firms have been studied so far, leaving aside service companies. It would be interesting to see whether service firms ask for different knowledge sharing catalysts compared to manufacturing companies, and whether the knowledge sharing mechanisms employed have a different impact on the innovation capability of firms. The research could be extended to encompass low-tech companies too.

Finally, it should be verified whether the results obtained can be generalized in other geographical settings, or whether substantial differences exist according to geographical location.

REFERENCES

Allee, V. (2003). *The future of knowledge: increasing prosperity through value networks*. Burlington, MA: Elsevier Science.

Almeida, P., Anupama, P., & Grant, R. M. (2003). Innovation and knowledge management: scanning, sourcing and integration. In Easterby-Smith, M., & Lyles, M. A. (Eds.), *Blackwell handbook of organizational learning and knowledge management* (pp. 356–371). Malden, MA: Blackwell Publishing.

Argyris, C. (1990). *Overcoming organizational defences*. Boston, MA: Allyn and Bacon.

Argyris, C., & Schön, D. (1978). *Organizational learning: a theory in action perspective*. Reading, MA: Addison-Wesley.

Argyris, C., & Schön, D. (1996). *Organizational learning II: theory, method and practice*. Reading, MA: Addison-Wesley.

Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to causal modeling: personal computer adoption and use as an illustration. *Technological Studies,* special issue on Research Methodology, *2*(2), 285-309.

Berger, P., & Luckmann, T. (1966). *The social construction of reality: a treatise on the sociology of knowledge*. New York, NY: Doubleday.

Brooking, A. (1996). *Intellectual capital: core assets for the third millennium enterprise*. London: Thompson Business Press.

Cantwell, J. (2005). Innovation and competitiveness. In Fagerberg, J., Mowery, D. C., & Nelson, R. R. (Eds.), *The Oxford handbook on innovation* (pp. 543–567). New York, NY: Oxford University Press.

Chin, W. W. (2003). A permutation procedure for multi-group comparison on PLS models. In Vilares, M., Tenenhaus, M., Coelho, P., Exposito, V., & Morineau, A. (Eds.), *PLS and Related Methods: Proceedings of the PLS'03 International Symposium.*

Chin, W. W., & Frye, T. (2003). *PLS-Graph version 3.00. Build 1017*. Texas: University of Houston.

Dalkir, K. (2005). *Knowledge management in theory and practice*. Oxford, UK: Elsevier Inc.

Davenport, T. H. (2007). Information technologies for knowledge management. In Ichijo, K., & Nonaka, I. (Eds.), *Knowledge creation and management: new challenges for managers* (pp. 97–117). New York, NY: Oxford University Press.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: how organizations manage what they know*. Cambridge, MA: Harvard Business School Press.

Drucker, P. (1988). The coming of the new organization. *Harvard Business Review*, (January-February): 45–53.

Edvinsson, L., & Malone, M. S. (1997). *Intellectual capital: realizing your company's true value by finding its hidden brainpower*. New York, NY: Harper Collins Publishers, Inc.

Fischer, M. M. (2001). Innovation, knowledge creation and systems of innovation. *The Annals of Regional Science*, *35*(2), 199–216. doi:10.1007/s001680000034

Friedman, V. J., Lipshitz, R., & Overmeer, W. (2003). Creating conditions for organizational learning. In Dierkes, M., Berthoin, A., Child, J., & Nonaka, I. (Eds.), *Handbook of organizational learning & knowledge* (pp. 757–774). New York, NY: Oxford University Press.

Kalla, H. K. (2005). Integrated internal communications: a multidisciplinary perspective. *Corporate Communications: An International Journal*, *10*(4), 302–314. doi:10.1108/13563280510630106

Maznevski, M., & Athanassiou, N. (2007). Bringing the outside in: learning and knowledge management through external networks. In Ichijo, K., & Nonaka, I. (Eds.), *Knowledge creation and management: new challenges for managers* (pp. 69–82). New York, NY: Oxford University Press.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242–266. doi:10.2307/259373

Nenonen, S. (2004). Analyzing the intangible benefits of work space. *Facilities*, *22*(9-10), 233–239. doi:10.1108/02632770410555940

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review*, *69*(6), 96–104.

Nonaka, I., Reinmoeller, P., & Senoo, D. (1998). The art of knowledge: systems to capitalize on market knowledge. *European Management Journal*, *16*(6), 673–684. doi:10.1016/S0263-2373(98)00044-9

Nonaka, I., Schamer, O., & Toyama, R. (2001). Building ba to enhance knowledge creation: an innovation at large firms. Retrieved from www.dialogonleadership.org/Nonaka_et_al.html

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. New York, NY: Oxford University Press.

Nonaka, I., Toyama, R., & Byosière, P. (2003). A theory of organizational knowledge creation: understanding the dynamic process of creating knowledge. In Dierkes, M., Berthoin, A., Child, J., & Nonaka, I. (Eds.), *Handbook of organizational learning & knowledge* (pp. 491–517). New York, NY: Oxford University Press.

Nonaka, I., Toyama, R., & Hirata, T. (2008). *Managing knowledge flow: a process theory of the knowledge-based firm*. New York, NY: Palgrave Macmillan.

Nonaka, I., von Krogh, G., & Voelpel, S. (2006). Organizational knowledge creation theory: evolutionary paths and future advances. *Organization Studies*, *27*(8), 1179–1208. doi:10.1177/0170840606066312

Polanyi, M. (1966). *The tacit dimension*. London, UK: Routledge & Kegan Paul.

Raelin, J. (2000). *Work-based learning: the new frontier of management development*. Reading, MA: Addison-Wesley.

Revans, R. (1983). *ABC of action learning*. Bromley, UK: Chartwell-Bratt.

Saint-Onge, H., & Wallace, D. (2003). *Leveraging communities of practice for strategic advantage*. Burlington, MA: Elsevier.

Shapiro, C., & Varian, H. (1998). *Information rules: a strategic guide to the network economy*. Boston, MA: Harvard Business School Press.

Stewart, T. A. (1997). *Intellectual capital: the new wealth of organizations*. New York, NY: Doubleday/Currency.

Subramaniam, M., & Youndt, M. A. (2005). The influence of intellectual capital on the types of innovative capabilities. *Academy of Management Journal*, *48*(3), 450–463.

Sveiby, K. E. (1997). *The new organizational wealth: managing and measuring knowledge-based assets*. San Francisco, CA: Berrett-Koehler Publishers, Inc.

Swan, J., Robertson, M., & Newell, S. (2001). Knowledge management: the next fad to forget people? In Barnes, S. (Ed.), *Knowledge management systems: theory and practice*. London, UK: International Thompson Business Press.

Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice: a guide to managing knowledge*. Cambridge, MA: Harvard Business School Press.

Wiig, K. (2004). *People-focused knowledge management*. Oxford, UK: Elsevier Inc.

Wold, H. (1985). Partial least squares. In Kotz, S., & Johnson, N. L. (Eds.), *Encyclopedia of Statistical Sciences* (pp. 581–591). New York, NY: Wiley.

KEY TERMS AND DEFINITIONS

Ba: It refers to the physical and/or mental space where knowledge generation and exchange take place.

Communities of Practice (Cop's): "Groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their understanding and knowledge of this area by interacting on an ongoing basis" (Wenger *et alter*, 2002, p. 4).

ICT-Based Knowledge Sharing: It refers to the processes of knowledge exchange which take place through the use of information and communication technologies (ICT).

Knowledge Sharing: It refers to knowledge exchange processes which take place both among people from the same organization and from different ones.

Knowledge Sharing Enablers: It refers to the set of organizational conditions (organizational design, culture, technological infrastructure and management systems in place) which facilitate the exchange of knowledge (i.e. knowledge sharing).

Knowledge Management: It refers to the management of knowledge acquisition, creation, transmission, storage and retrieval processes.

People-Based Knowledge Sharing: It refers to knowledge exchange processes which take place by means of social/direct interaction among people.

This work was previously published in Handbook of Research on Communities of Practice for Organizational Management and Networking: Methodologies for Competitive Advantage, edited by Olga Rivera Hernáez and Eduardo Bueno Campos, pp. 185-200, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 5
Integrating 'Designerly' Ways with Engineering Science:
A Catalyst for Change within Product Design and Development

Ian de Vere
Swinburne University of Technology, Australia

Gavin Melles
Swinburne University of Technology, Australia

ABSTRACT

The fields of design and engineering both contribute to product design and development. Increasingly design teams require an integrated approach in environments where mutual understanding and respect replace traditional professional rivalries. These new synergies both enhance communication and understanding between designers and engineers and lead engineering into new areas of professional activity. Engineers are integral to the product development process, but change in product development and manufacturing requires new responsibilities; design engineers must assume a greater role to achieve successful product realisation. However, to be effective engineers must develop new skills; creative design ability, understanding of societal and environmental impacts and a human-centred approach. These themes, not typically addressed by engineering curricula are evident in a new approach to engineering education - product design engineering. This chapter addresses issues confronting product design and development and examines the emergence of this new engineering professional in response.

DOI: 10.4018/978-1-4666-1945-6.ch005

INTRODUCTION

Product design and development (PDD) responsibilities have changed. There is greater focus on sustainable design, socially responsible design and design for need. Opportunities exist for design teams to make a positive commitment to the welfare of global communities whilst advancing technologies that support sustainable development. It is no longer appropriate for designers and engineers to serve solely the interests of business; instead PDD teams must understand the potential for design to make a greater contribution to lives and society.

As the roles and responsibilities of product design and development teams are reformed, so too are their professional composition. The single discipline purity of the traditional industrial design consultancy has evolved into an interdisciplinary team, where designers and design engineers collaborate harmoniously to provide an extended palette of services. Product design teams require an integrated and collaborative approach in environments of understanding and mutual appreciation. The product design and development process is enhanced by these new synergies between engineers and designers, as is the progression of the engineering designer into new areas of professional activity.

Whilst design engineers have always been an integral part of the product development process, their roles have traditionally been confined to working within constraints and defined parameters to achieve closure to the product realisation stage. However, emerging trends in manufacturing and revised professional responsibilities require design engineers to have a greater role in product design and development, particularly in the conceptual design and embodiment stages. Yet to be effective, they require an extensive palette of new skills; creative design ability, a thorough understanding of the societal and environmental impacts of their professional activities and a human-centred and responsible approach. These attributes are not characteristic outcomes of traditional engineering curricula, but are evident in product design engineering courses.

This new engineering discipline results from the integration of two traditionally disparate professions; mechanical engineering and industrial design. It responds to the need for interdisciplinary professionals and a greater participation in design teams by engineers conversant, indeed accomplished, in the product design and development process. These new engineering pedagogies support the changing role of the engineering designer and are catalysts for significant change in product design and development through greater team synergy, interdisciplinary understanding and communication. "Times of great flux call for those who can cross disciplines, who can see and understand the big picture." (Akay, 2003)

BACKGROUND

It has been observed that the boundaries between the design and engineering can inhibit both innovation and successful product realisation, particularly in the product design and development milieu. "These two mindsets often clash as one seeks to broaden the scope of the problem, while the other is working to achieve closure." (Fry, 2006)

In Engineering Design Methods: Strategies for Product Design, Nigel Cross notes that "the increasing competition for consumer markets and the growing awareness of the importance of design for the market has led to reinforcement of the view that successful design can only be accomplished by an integration of the skills of both engineering and industrial designers." (Cross, 2000)

This trend is evident both in the traditional industrial design consultancy and in the manufacturing sector where there is increased demand for engineers who can operate effectively in a variety of environments within global multidisciplinary teams. Engineers, particularly those in product design and development, are now expected to creative, flexible and adaptable, responsible and human-centred designers. "In this evolving world, a new kind of engineer is needed, one who can think broadly across disciplines and consider the human dimensions that are at the heart of every design challenge." (Grasso & Martinelli, 2007)

Conflict between disciplines has been common in product design and development, where differing professional approaches can destabilise the progression of design ideals. A long-standing cultural conflict exists between designers and engineers, one that can be traced to the foundation of their approach to problem solving and design. Fry (2006) defines industrial designers as "looking for new contexts and opportunities for innovation" whilst engineers "predominately work to define a set of parameters and target values up front that would define a specific, successful solution within a narrow range."

It is important to acknowledge that the approach of engineers to design activity results from the attitudinal emphasis of their education. Many engineering curricula offer limited exposure to the design process and often graduates are inexperienced in open-ended problem solving and fearful of the uncertainty of the design process. Engineers trained only in the 'science' and 'methods' of engineering are often hesitant to engage in intuitive or creative processes that may lead to an unexpected outcome.

This absence of cohesion between engineers and designers can lead to misunderstanding, a lack of appreciation, and subsequently, professional distrust and discord within the PDD team. Professional disparity between engineers and designers can result in inefficiencies in the PDD process and can be detrimental to successful product realisation. To facilitate developments in product design and development, the disparities between the professions must be addressed through engineering educational programs that seek to emphasise design, encourage creativity and innovation, resulting in improved cross-disciplinary communication, understanding and mutual appreciation.

We need "a new generation of adaptable, flexible, well rounded and innovative professionals" (Stouffer, Russell, & Oliva, 2004). However innovation and creativity does not readily emerge from typical engineering process. To be creative one needs to seek the unexpected through a process of divergent thinking. Unfortunately engineering design and industrial design follow differing methods; "convergence is at the core of the engineering process...divergence is at the core of the industrial design process" (Fry, 2006) It is therefore apparent that to develop creative engineers, their educational experience must not only include rational applications of science, but also experience in open-ended problem solving and familiarity with what Cross describes as 'designerly ways'. (Cross, 2001)

'Designerly ways' suggests a way of thinking that is common in designers, but which differs significantly from those with a techno-scientific approach, such as engineers. Cross observed that scientists problem-solve by analysis, whereas designers problem-solve by synthesis. This constructive and often intuitive approach uses non-verbal thought and communication, translates abstract requirements into concrete objects and contributes to the solution-focussed approach designers take to solving ill-defined problems. Engineering education can learn much from design pedagogy if engineers are to develop creative problem solving skills early through education, rather than later in the workplace.

The engineering profession must progress from the role of technical service provider, to a profession that leads change through understanding of the human, environmental, societal and cultural challenges and the consequences of professional activity.

ISSUES CONFRONTING PRODUCT DESIGN AND DEVELOPMENT

Design is a fundamental building block of innovation and a critical enabler of competitive advantage. Product design is essential in addressing the challenges that modern societies face in the 21st Century; whether it be enabling the delivery of technological, scientific and medical advances, mobilising public opinion, empowering communities, or simply improving the experience of a product. (Design Victoria, 2008)

However there are many challenges facing the product design and development industry as we move into the 21st century. These range from the decline of local manufacturing and the impact of a global economic downturn, to new opportunities afforded by emerging markets, the integration of new disciplines and the redefinition of existing professional responsibilities. Sustainable practice and socially responsible design will continue redefine the landscape and the PDD industry must take responsibility for its activities and show leadership as drivers of change and innovation.

The fortunes and growth potential of the product design industry are too closely allied to the manufacturing sector, an area of recent considerable upheaval and changing alliances. The transfer to offshore manufacturing has had unprecedented impact in the design sectors of many countries, as have the competitive forces of a global economy and the reduction in Government protection of local industries. Many product design environments have witnessed a rationalisation of product lines, a reduction of local content and a shift in outsourcing towards ‘design and supply’ contracts.

In response product designers must take responsibility for their destiny through business diversification, entrepreneurship and the creation of new opportunities through flexible and adaptable approaches. PDD must reduce its reliance on the manufacturing sector (as a sole source of revenue); in turn this will lead to greater diversification of services into new sectors, including interdisciplinary activities which are solution, rather than market focussed.

Socially responsible design initiatives and sustainable solutions are imperative and offer PDD teams a diversity of new opportunities. This will need to be complimented by design-led initiatives as PDD teams redefine and create new avenues of activity, and respond to societal needs on a global basis.

“A new economy is evolving and it needs products and services that have been designed to eliminate environmental impacts.” Kel Dummet, Sustainability Victoria (Design Victoria, 2008)

As the manufacturing industries adjust to a changing economic and social environment, new markets are emerging with diversified consumer behaviour and expectation. It is therefore essential that PDD teams develop interdisciplinary attributes that successfully integrate design, engineering, management and marketing in a flexible milieu of understanding and respect, with a high level of creativity and consideration for environmental, societal, cultural and economic demands.

Typically, poor product design decisions can adversely affect manufacturing costs, assembly schematics, quality control, product to market lead-times and overall product success. This can be addressed through consideration of manufacturing issues earlier in the PDD process, in particular design for assembly, with user-centred design addressing not only the end user, but also production, transportation, and where appropriate, installation personnel. This early integration of manufacturing consideration into the design process can be facilitated by the inclusion of design-skilled engineers into PDD, bringing production concern to the forefront of product embodiment decisions.

Product design is at a turning point, one which will redefine the industry and allow significant societal contribution to be made through new directions of activity. The key issues facing the PDD industry; sustainability, social responsibility, design for need, global distribution of design and manufacturing processes, new and emerging markets and shifting economic power will be the focus of the next generation of product designers, who must be sufficiently informed and skilled to lead the inevitable reform of industry practice.

In Australia, the Design Victoria report "Five Years On: Victoria's Design Sector 2003-2008", (through a survey of 340 design service providers and 1253 businesses) identifies design as making an essential contribution of economic growth, yet also identifies that graduates lack appropriate skills including business acumen and production and manufacturing knowledge. (Design Victoria, 2008) This trend is apparent in both the design and engineering sectors. Typically design curricula have become generic and holistic whilst engineering curricula continues to focus on the science rather than preparing graduates for engineering practice. Clearly the challenges faced by industry must be met firstly by the educational community through revised, industry relevant and progressive curricula: industry reform should be led by the next generation of designers and engineers.

So how do we define the roles and responsibilities of future PDD teams and what will be the challenges they face and the contributions they make? Firstly we must anticipate the impact of trends in product design and development, respond to the emergence of new technologies and constraints, and most importantly we must involve engineering education in the conversation. We cannot hope to foresee all of the challenges that our graduates will face throughout their careers; however we can prepare them for significant change through educational programs that deliver flexible, adaptable and responsible engineers who are lifelong learners and creative problem solvers.

Tertiary education should be the foundation of improved and redefined professional outcomes for product design. New paradigms such as product design engineering (discussed later in this chapter) are catalysts for significant change and directly respond to industry needs and societal demands. The integration of disparate disciplines (industrial design and mechanical engineering) facilitates greater professional synergies, and thus, efficiencies in the product design and development process. It is essential that all PDD team participants, particularly the engineers, are fully engaged and cognisant in all aspects of the product's 'function-aesthetic-experience' and demonstrate an eagerness for innovation. The challenges of future product design and development will be greatly facilitated through teams which include these 'integralists' (Eekels, 1987) who can articulate the overall product vision in an collaborative environment devoid of traditional professional rivalry.

The product design engineering discipline emerges from industry-informed curricula which develop well prepared, highly skilled and adaptable practitioners vital to the security and future competitiveness of the PDD industry.

DEFINING THE ROLE OF ENGINEERING IN PRODUCT DESIGN

What is the role of this new discipline, the product design engineer, in the PDD process? Before we address this, it is important to define the roles of engineers and industrial designers and examine disparate claims regarding who 'owns' the process.

Ulrich and Eppinger in their definitive text "Product Design and Development" (Ulrich & Eppinger, 2004), devote a chapter to Industrial Design (ID) where they define the role of ID in the PDD process. In addition, they endeavour to quantify the effectiveness and utilisation of industrial designers and the timing of their involvement in the development process. They state that "typically, ID is incorporated into the product development process during the later phases for a technology-driven product and throughout the entire product development process for a user driven product." They also declare that the ID process is a "subprocess of the product development process" and that "the technical nature of the problems that confront engineers in their design activities typically demands more development effort than do the issues considered by ID".

The reader may discern two things from the position taken by Ulrich and Eppinger. Firstly, that the product design process is not directed from a design perspective and secondly, that ID is a non-technical and potentially optional phase of PDD which is not integral to product development. With their statements, they are reinforcing a position not uncommon in production engineering circles, but one which devalues not only the contribution that industrial designers make to product design and development, but also the potential for designers to contribute significantly to product forecasting and planning.

With due respect to Ulrich and Eppinger, whose writings on PDD are exemplary and extremely valuable to students and practitioners seeking to develop a thorough and 'validate-able' PDD process, the industrial design profession may wish to present an alternate stance.

The early involvement of designers greatly informs the product design and development process and can actually drive technology development through earlier consideration of user needs and more appropriate and targeted product specification. Indeed it is essential that the early stages of all projects (regardless of whether it is for user or technology led products), involve those personnel with a human-centred focus and a divergent thinking approach.

It has been our professional experience that all too frequently Industrial Design input is integrated far too late in the project, often at a convergent stage of development, where the product's technical specification has been established. This limits the scope and potential of the product and forces the industrial designers to adapt and package often unsuitable or inappropriate technology and components that have been developed for poorly defined user needs. The impact of the designer's understanding of market, expertise in socially responsible design (SRD), sustainable design, ergonomics, user needs, product interaction, aesthetics and product differentiation combined with their intrinsic quest for creative solutions and innovation cannot be underestimated in product scoping and planning.

Accordingly, design practitioners must be central to the product planning process, not introduced retrospectively when parameters are pre-defined and existing constraints restrict innovation. This can be addressed in two ways, with the establishment of a collaborative team comprising both industrial designers and engineers early in the product planning stages, or with the inclusion of a new type of engineering professional; the product design engineer.

These new engineering professionals though the integration of 'designerly ways' into the practice of engineering are well placed to drive a user-centred and creative design philosophy in the planning stages of PDD, particularly in technology driven products where typically industrial designers may not be included in the discussion.

New Directions for Engineering

It is apparent that as product design and development evolves to meet new challenges that its engineers will require:

- improved design emphasis (including sketching ability),
- a creative approach,
- a socially responsible design philosophy,
- an emphasis on sustainable product design and manufacture and
- advanced open-ended problem solving ability

An Emphasis on Design: Integration of Designerly Ways

Most engineering faculties have been reticent to incorporate any significant design course components, at least until the final or capstone years. Faculties prefer to build a solid foundation of engineering science, but without supporting design activities that provide opportunities for this knowledge to be applied. "We do not teach the language of design, preferring instead the language of mathematics". (Dym, 1999) This view is supported by a perusal of typical engineering curricula that finds that negligible design projects and a notable absence of divergent thinking or creativity agenda.

Design has always been fundamental to the practice of engineering, and thus should be a key constituent in engineering education. Unless design and creative skills are a focus of learning activity, graduate engineers will continue to be technically competent, but will not prepared for the practice of engineering or the challenges they will face professionally. Both Fox and Cross have identified that there is 'an educational justification for design' as a means to develop cognitive skills and real-world problem solving abilities. (Fox, 1981) (Cross, 2001)

Yet it appears most engineering curricula continue to focus on the theory-based science model, denying students the learning opportunities afforded by design projects, particularly in the application of engineering science to real world practice. Design needs to move from the periphery to a central role in engineering education if we are to graduate innovative and adaptable engineers, and enable more valuable engineering input in the PDD process.

"One of the consequences of design-focused education is that students learn that they are applying knowledge in differing forms to serve different ends, which means that they can become fluent translators of engineering languages." (Dym, 1999)

Fostering Creativity

It appears that many in engineering faculties do not value creativity or do not readily acknowledge its importance in engineering practice. This despite many calls for a focus on fostering creativity. "Creativity is the essence of engineering. Yet creativity is neither explicitly taught nor promoted in the engineering curriculum." (Santamarina, 2002)

Creativity is frequently misunderstood, and is often viewed with an element of mistrust, especially by those outside the design community. This is due to creativity being viewed as resulting from 'ex nihlo' (something from nothing), rather than from less confronting processes involving structured problems and expected or predefined solutions. However teaching engineers that creativity involves "a non-linear, unstructured and flexible approach to solving problems and generating ideas" (Pappas, 2002) can facilitate acceptance and willingness. Engineering students must become comfortable with the creative process, if they are to embrace unexpected outcomes and pursue innovative solutions. "Making the strange familiar – accepting creativity as a desirable mindset and attribute of engineers – is a tangible and realisable goal." (Stouffer et al., 2004)

So what is creativity and why is it important in an engineering educational context?

Creativity emerges from a divergent thinking process that allows the designer unconstrained exploration, the use of intuition and reflection to respond to a problem with a solution-focussed approach. Creativity seeks to generate new, unique and unexpected solutions.

"Creativity is of paramount importance in engineering for it endows one with insight and discipline to seek out and address problems from the boundaries of different engineering disciplines" (Ghosh, 1993)

Creativity involves having unusual ideas, tolerating the unconventional and seeing unexpected implications. (Cropley & Cropley, 2000) This can be rather challenging for engineering students who are more comfortable working within defined parameters and tackling constrained (rather than open-ended) problems, and who typically, have a tendency to fixate on prior solutions. "Students must be aware that instruction in creative thinking will not provide the certainty offered them in most engineering tasks. Creative thinking is a more ambiguous endeavour than most engineers are used to or skilled in...there are no right answers" (Pappas, 2002).

Yet without a creative methodology, the potential for engineers to contribute to the PDD process is limited and innovative engineering solutions are less likely. Creativity results from a flexible and open approach to problem solving, an approach dependent on confidence and a willingness to take risks and trust intuitive ideas. Typically, these attributes have not been developed through the engineer's education, but have been gained through extensive industry experience. To better prepare graduates for the demand of industry and to enable more innovative engineering practice, we must foster creativity in students. To achieve this, curricula must facilitate significant experience of design problems, in an environment that appreciates unexpected solutions, tolerates failure and nurtures students through the framing and resolution of ill-defined problems.

Initial attempts at fostering creativity focussed on teaching specific techniques, for example Osborn's brainstorming. However this ignored the non cognitive aspects of creativity including motivation and self confidence (Cropley & Cropley, 2000), which are integral to student engagement in the creative process. Engineering education must focus on developing problem solving skills, in particular, to reduce anxiety about unexpected solutions, develop confidence in the student's creative potential, and instil a desire for seeking creative and innovative solutions. It is no longer sufficient for students to be engaged in simple problem solving activities; experience in the 'practice of engineering' must occur through experiential learning projects that allow science to be applied to real world problems.

The Importance of Sketching

There is significant research that links sketching to creativity, yet many engineering curricula fail to develop the sketching skills essential for design conceptualisation and communication and resultantly limit the ability of engineering graduates to explore design possibilities. Sketching enables abstract idea development and is not just a documentation and communication process; rather it facilitates the creative process through contextual citing of the design intent. Without sketching ability, many engineers struggle to find the unintended consequences that stimulate the design exploratory process in what Schon and Wiggins (1992) called the 'reflective conversation with the situation'.

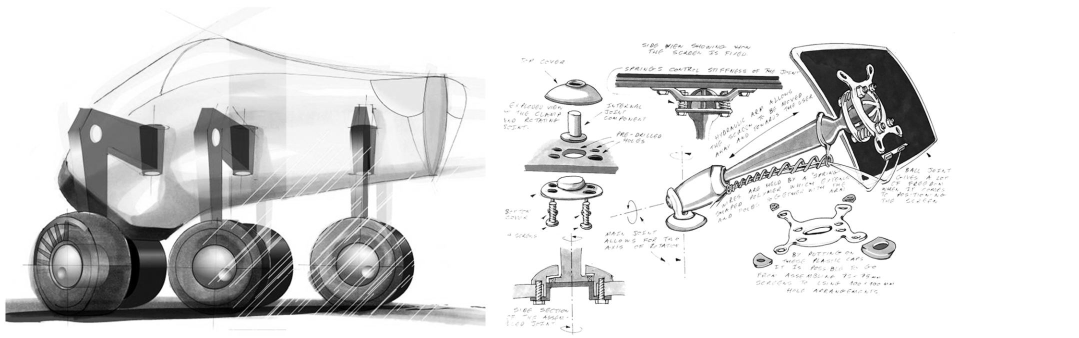
"It appears that the very design process is limited by the ability to use graphics as a cognitive extension. This implies the need for training... not only in the standard drafting skills, but additionally in the ability to represent concepts that are more abstract and best represented as sketches" (Ullman, Wood, & Craig, 1990)

Sketching is fluent, flexible and inaccurate and because sketches are 'rough' there is less reluctance to discard them in favour of alternative designs. Sketches are fast, implicit, inexact and abstract, require minimal commitment and serve both analysis and simulation requirements. (Lipson & Shpitalni, 2000) Goldschmidt in his definition of sketching dialectics, refers to a dialogue between reflective criticism ('seeing that') and analogical reasoning and reinterpretation of the sketch ('seeing as'). (Goldschmidt, 1991)

Most importantly, sketches allow the designer and engineer the opportunity to explore as many concepts as possible before moving into the detailed design stage. That sketching is integral to the creativity process, is reinforced by studies of industrial design students which compared skilled and unskilled sketchers (Verstijnen & Hennessey, 1998) who found that those with sketching ability were advantaged by externalisation of mental imagery.

Yet many engineering faculties rely solely on CAD for design, neglecting the possibilities that sketching offer in developing creative solutions. As sketching facilitates "the creative shift to

Figure 1. Differing roles of conceptual sketching: Exploration of form (left) and abstraction of technical details (right)



new alternatives" (Goel, 1995) it is imperative that engineering students are competent sketchers, capable of articulating unformed ideas and design intent.

SOCIALLY RESPONSIBILITY DESIGN (SRD)

It is imperative that engineering serve the community in a socially responsible, sustainable and culturally sensitive manner. "Because of the intrinsic connection of engineering design with values, the engineer as designer shall not only be answerable for his/her engineering capabilities, but also and always for his/her ethical conceptions and behaviour as a moral person." (Eekels, 1994) Not only do product design teams need to be aware of the consequences of their professional actions, they must also be cognisant of the opportunities afforded them to lead positive change.

Designing for our complex societies requires anticipation of future needs and cultural sensitivity, but there exists a greater need, that of those at the base of the pyramid (the other 90%). Product design teams must address the societal needs of developing nations through designs that are indeed life supporting, sustainable and empowering. Urgent action is required in many critical areas including clean drinking water, sanitation, renewable energy, and healthcare and disease prevention.

A new engineering conscience is needed, one that emboldens 'design for need' over the material needs of first world consumers and repositions product design and development not as a functionary of business, but as the leader of change for greater social good. Engineering education is well positioned to make a significant contribution to determining attitudinal change amongst the student cohort and thus developing social conscience and sustainable and ethical practice in its graduates. The next generation of engineers can contribute to appropriate product design and development through responsible, ethical and culturally and environmentally sensitive design practice.

The engineering profession must progress from the role of technical service provider, to a profession that leads change through understanding of the human, environmental, societal and cultural challenges and the consequences of professional activity. This can only be achieved through a systematic and thorough grounding in the principles of socially responsible design (SRD) combined with opportunities to apply sound engineering practice to real world problems throughout the educational journey.

Sustainability

The UN has estimated that 80 percent of all product-related environmental impacts are determined during the product design stage. If issues of resource depletion, energy usage, material reuse and recycling, environmental degradation and climate change are to be addressed in sufficient time, environmental and social considerations need to be integrated early into the education of design engineers. Our responsibility as educators cannot be just to our students; we must be answerable to societies in general for the consequences of our graduates' professional behaviour. Our engineering graduates need not just awareness of the issues, but an embedded ethical philosophy and the tools to effect reform through sustainable design and product engineering.

The decade 2005-2014 has been declared the United Nations 'Decade of Education in Sustainable Development'. This initiative aims to "encourage changes in behaviour that will create a more sustainable future in terms of environmental integrity, economic viability and a just society for present and future generations." We must respond to this call through a targeted program that teaches not only awareness, but also the tools to affect behavioural change and to lead reform in design practice and manufacturing. The next generation of product designers and design engineers must direct initiatives to address the challenges faced by the manufacturing sector. Energy and material consumption and carbon emissions must be reduced, environmental degradation must be halted and alternative technologies developed. But most importantly the needs and aspirations of those currently outside the first world consumer societies must be addressed; not only are they the unwilling participants in relentless and unsustainable development, they also lack the basic elements that constitute a safe, healthy and equitable existence.

"Sustainability cannot be conducted on the sidelines. It can only be achieved through a paradigm shift which results in sustainability becoming part of everyday life, directing the way in which communities and individuals make decisions that contribute to the development of broad social goals." (Hammer, 2007) Engineering and design educators are ideally placed to make a positive contribution to reducing the impact of product manufacture and life cycle through the development of pedagogy that facilitates this paradigm shift.

Many of these goals can be addressed through innovative curricula that inform, encourage and require design project outcomes that include sustainable manufacturing processes and material selection, life cycle analysis (cradle to grave), design for assembly and disassembly and a holistic approach to product design that questions not only the impact of the product but also the 'need' for the product to exist at all. Designers and engineers must be taught to question established practice and to be sufficiently confident to make strategic calls that may defy established practice in search of a more sustainable product solution. "Designers have responsibility over the choices they make in design processes." (Papanek, 1985) Sustainable design practice must merge as an entrenched design methodology that is at the core of all professional activity. This will require attitudinal change and will require product design teams to serve not only their immediate stakeholders (clients and customers), but also the needs of global societies, including those outside their own market.

Sustainable design addresses the 'triple bottom line', a reporting framework that considers ecological and social aspects, in addition to the traditional financial measures. These goals demand that corporate responsibility be to its stakeholders not just its shareholders; consequently anyone who is influenced or affected, either directly or indirectly, must be considered. Appropriate design, fair trade initiatives, reduced ecological footprints and true accountability must be paramount in the minds of product design teams if we are to reduce the negative impact of our activities in product design and manufacturing. Victor Papenek famously declared that "there are professions more harmful than industrial design, but few of them." (Papanek, 1985) In response, product design and development teams must take the initiative to secure a sustainable lifestyle for future generations through well considered and responsible design and engineering practice; the role of engineering education is to provide the knowledge, aptitude and attitude required to lead this paradigm shift.

Problem Solving: 'Wicked' and Open-Ended Problems

It is evident that the 21st century design engineer will be confronted with ill-defined problems; design problems that will not be solved solely through the appliance of engineering science. These 'wicked' problems (Rittel & Webber, 1973) will require a new engineering approach, one that incorporates divergent thinking, creative and intuitive processes, and a willingness to embrace unexpected solutions; a designers approach to problem solving.

As discussed earlier, design is a divergent activity, whereas engineering thinking is mostly convergent. It is no surprise that mechanical engineering students struggle with open-ended problem solving when compared with design students. Most engineers are more comfortable with constrained problems, where the parameters are set and the outcome is, if not immediately apparent, determined by constraints afforded by material and processes. By contrast designers 'live' for the creative moment: they thrive on the pressure of uncertainty, exult in the quest for uniqueness and triumph with innovative solutions.

Design is an intensely personal pursuit where one makes a commitment to the problem that involves significant personal vulnerability and a willingness to face uncertainty and failure head-on. It has been said that a designer is only as good as his/her last design, as such designers must constantly expose themself to intense scrutiny as their outcomes are measured against unforgiving criteria, and judged by a market often driven purely by economic rationalism. Successful design practice requires a degree of self examination and confidence in the face of critical scrutiny that is not required by other professions; in design mere competency is insufficient.

However, many of the problems facing design engineers in product design environments will also be poorly defined and their first challenge will be problem definition; this will need experience and confidence when faced with uncertainty. Students need a learning process of construction and confrontation rather than memorisation. Project-based learning allows students to define and analyse the problem, develop alternative strategies to problem solving and to build, enhance and practice their expertise. But not all project-based learning will afford intense exploration and challenge students to work without constraints or parameters, as will an open-ended problem.

Open ended design problems "force students to think creatively and ultimately foster in them an appreciation for developing creative solutions." (Ghosh, 1993) This requires students to tolerate the unusual, unconventional and unexpected, and to become comfortable with divergent thinking processes. Open-ended problems require problem finding, evaluation, definition and resolution skills together with ideational fluency. They take students away from a problem 'solving' role into a position of exploration and synthesis that enforces a more holistic reasoning approach with critical analysis and creative expression. Experience with ill-defined design problems (that are not amenable to the techniques of science and engineering), is invaluable in engineering education. Students must develop flexible and divergent dichotomies and eliminate any tendency to fixate on prior solutions; this will require a new approach to the teaching of engineering design.

To address ill-defined problems, engineers must be confident seeking solutions outside their traditional fields of expertise, and comfortable using intuition to solve problems, rather than relying on solely on mathematical and scientific solutions.

The Emergence of a New Engineering Discipline within PDD

Industry trends indicate that role and responsibilities of the engineering designer in product design and development have changed from technical resolution to a more central role in the planning and design process. Design engineers are often engaged in creative 'front end' design activities (typically the domain of industrial designers) in addition to systems engineering and manufacturing resolution roles. "Current trends in technology and our increasingly complex society and workplace require engineers to have a wider variety of skills and a broader understanding of engineering as a discipline if they are to be successful." (Pappas, 2002)

Unfortunately it appears that traditional engineering curricula has not responded to the repositioning of engineering in PDD environments, and continues to focus on the 'theory' rather than the 'practice' of engineering, neglecting the development of design and creativity in students. "Historically, engineering curricula have been based largely on an 'engineering science' model." (Dym, et al, 2005) However, curriculum developments, particularly during the last decade, have led to the emergence of new engineering pedagogy that support integrated design environments. These new educational models show evidence of an attitudinal shift that leans "toward a more explicit recognition of design as a distinguishing feature of engineering practice and a motivating factor in the learning of engineering." (Dym, 1999)

Product design engineering (PDE) has created a new style of engineering designer; creative, flexible and adaptable, sustainable, ethical and responsible. Product design engineering whilst a relatively new engineering discipline, is gaining support with global distribution of curricula. These pioneering curricula integrate the traditionally disparate disciplines of industrial design and mechanical engineering to develop a creative 'next generation' engineer more adapt at the changing roles of product design and development.

Swinburne University of Technology PDE graduates have already made a significant impact in local product design and development. From observation of (and discussions with) leading industrial design consultancies in Melbourne, Australia, it is apparent that the product design engineering model is leading a trend in product design and development team constitution, with these interdisciplinary professionals defining a new approach to product design and breaking down traditional professional rivalries.

All of the major consultancies (and most companies in the manufacturing sector) are employing PDE graduates as they provide extended services to their clients and engage in brand building, global manufacturing and sustainable design solutions. Upon investigation, it appears that in the three leading industrial design consultancies examined closely, Swinburne product design engineers represented approximately 20-30 percent of the design staff, the balance being industrial designers. This is a significant achievement for a single course that has only seven years of graduates and that is lacking in public awareness as a distinct design discipline.

Ensuing conversations revealed that when employing a graduate, consultancy directors are more inclined towards employing a product design engineer for a traditional ID role, on the basis of their greater technical ability. It is felt that product design engineers 'reflected the industrial design graduates of previous generations', were more industry ready and could make a valuable contribution immediately upon employment. By contrast most ID graduates lacked technical ability and an understanding of professional practice, as a result of an education process less rigorous and more generic in nature. In consultancies whose activities were defined as approximately 20 percent front end creative, 60 percent detail design and 20 percent project management, the PDE's were a better investment with several consultancy directors declaring that they prefer product design engineers when hiring junior staff.

It was said that ID graduates were still slightly better at purely conceptual front-end design and that at present the most preferred employee was an industrial designer with 5-10 years industry experience; although this may change as PDE graduates reach professional maturity.

CASE STUDY—THE SWINBURNE PRODUCT DESIGN ENGINEERING CURRICULA: A CATALYST FOR CHANGE

Overview

Product design engineering represents a different approach to the teaching of engineering. It is an exemplary example of interdisciplinary education that not only develops a creative and human centred engineer, but understands the need for a more integrated approach to product embodiment. Without neglecting the science of engineering, it seeks to develop graduates who utilise the intuitive creativity of the design practitioner, united with sound engineering design method, to create a new approach to product design and development.

This educational paradigm is evident across a range of educational institutions worldwide as engineering faculties seek to address shortcomings in traditional curricula; in particular the notable absence of embedded design activities and negligible fostering of creativity. There are many models emerging either through targeted product design activities in mechanical engineering curricula (e.g. the product design major at Stanford), double degree courses or fully integrated programs such as the PDE courses offered in the United Kingdom, Europe and Australia. But all of these courses address the identification by Dym, Pappas, Akay, Santamarina and others, that engineering education needs reform if we are to develop engineers who are 'creative thinkers'.

It is important to clarify that the product design engineering paradigm discussed here is the product of fully accredited engineering courses with an engineering qualification (e.g. BEng, MEng) outcome and should not be confused with variations of the industrial design paradigm; PDE graduates are a new category of engineering designer.

However this new engineering curricula is not the only educational response to develop greater synergies between engineering and industrial design. The industrial design engineering (IDE) programs prevalent in Europe and the UK respond to the same industry requirements but through an expansion of industrial design curricula to incorporate engineering content.

Product design engineers are capable of operating with distinction in both professions, they utilise the divergent, reflective approach of designers, but engineer with the disciplined thoroughness and scientific method of engineers. More importantly, their integrated approach to product design incorporates a level of engineering proficiency not evident in industrial design, and their engineering methods benefit from the creative, flexible and human centred approach of the designer. This allows a valuable contribution to the PDD environment as they design from the 'inside out'; firstly engineering the functionality and then designing the user experience.

Product design engineers with their interdisciplinary attributes utilise intuitive creativity with sound engineering design method, to develop innovative solutions to complex, ill-defined problems. This results in a more proactive role in product design and development as these new engineers engage in initial problem framing activities, share responsibilities with industrial designers, and produce user-centred design solutions.

Background to PDE

Product Design Engineering originated as a new engineering discipline in Glasgow, Scotland in the late 1980s as a changing manufacturing environment demanded a professional fluent in both engineering and design. The first product design engineering course resulted from collaboration between Glasgow School of Art, (Industrial Design at the School of Design) and the University of Glasgow, Department of Mechanical Engineering and was soon followed by similar pedagogy at the University of Strathclyde. It intended to alleviate the gulf between the design and engineering professions and address disparate technical and cultural issues. It also aimed to introduce many of the educational topics raised by Donald Schön in 'Educating the Reflective Practitioner' including reflection-in-action and joint experimentation on open-ended problems. This new engineering curriculum integrated the design and engineering disciplines to create one model of reflective practicum (Green & Kennedy, 2001) with "design a continuous thread running through the teaching", an attribute preferred by the Institution of Mechanical Engineers, in the Grant Report, The Formation of Mechanical Engineers: Present and Future Needs.(IME, 1985)0

The importance of this interdisciplinary educational approach is evidenced by the emergence of similar pedagogy within the last decade, particularly in the United Kingdom and Western Europe (PDE is now offered by least twenty-five Universities worldwide), and the success of the Swinburne University of Technology PDE program in Melbourne, Australia, established in 1997. As industry enthusiastically embrace these 'new' engineering graduates, these innovative curricula appear to be feasible alternatives to traditional methods for teaching engineering, particularly in areas where design skills, creativity and problem solving ability are valued professionally.

The intent of the product design engineering curriculum is best described by a sample of course mission statements from institutions offering this new curriculum, as follows:

"There are increasing pressures, both from existing and emerging world marketplaces, for products which not only respond to the needs of function, user and society, but which can be brought to market ever more rapidly through state-of-the-art development and manufacturing processes. Industries which develop, manufacture and market today's products need high calibre graduates equipped to handle these processes with management skill and creative drive, and the PDE programme develops graduates with these skills." (Glasgow School of Art, 2009)

"The Product Design major concerns itself with the conception and design of products, services and experiences for the benefit of society. The program teaches a design process that encourages creativity, craftsmanship and personal expression and emphasises brainstorming and need finding to discover latent or un-served human need." (Stanford University, 2009)

"This course bridges the worlds of design and engineering, producing true hybrid professionals who will move immediately into companies as design engineers and product development specialists. This rare degree format has proved to be highly regarded by industries worldwide for producing valuable employees." (Brunel University, 2009)

From these course descriptors, it is apparent that, regardless of the country or institution, the product design engineering paradigm intends to develop a creative and human-centred design approach to the practice of engineering within the PDD environment.

This new engineering curricula is dependent on a number of distinct and crucial elements which are integrated throughout all areas of the learning process. These include:

- interdisciplinary education and skill integration
- an emphasis on design – integration of designerly ways
- the fostering of creativity
- sketching
- social responsible design
- sustainability

- experience of 'wicked' and open-ended problems
- experiential project based learning

Project Based Learning: Learning from Design

"Innovation requires creativity. The easiest vehicle for promoting creativity and for developing the student's decision making ability is the design project." (Eekels, 1987)

Developing product design aptitude requires experience; opportunities to learn, to explore, to experiment and to fail. Traditional engineering pedagogy, reliant on a theory-based scientific model does not offer many such learning opportunities. Students are not required to apply their newfound knowledge to the design of product solutions for real world problems; it is more likely that students will be tested on their science through examination and through solution-focused problem solving activities that lead to predetermined outcomes. This leads not to creative and innovative engineering practitioners, but rather, to pragmatic and inflexible professionals who readily fixate on the security of known solutions.

On the other hand, a design curriculum is 'learning in action' where students learn theory, and develop creativity and design experience through project-based or experiential learning processes. Students are encouraged, indeed required, to explore the limits of their imagination, to push the boundaries of technology, processes and materials: to investigate the possibilities and extend their own potential. Project based learning affords opportunities for the application of the known, the science of engineering to be applied to the unknown, often resulting in an unexpected solution. The joy of teaching design projects is that every brief will result in unique and innovative designs from each and every student; there are no correct or incorrect answers, only creative and unexpected solutions.

Project based learning is constructive, participatory and problem driven; an exemplary pedagogical model of 'learning by doing' and the most effective way to educate creative engineers for product design and development.

Creativity and Innovation

Felder (Felder, 1987) notes that creativity can be fostered through exercises that focus on:

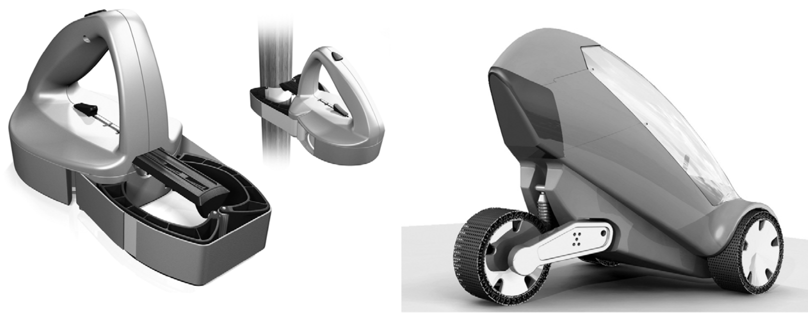
- ideational fluency
- poorly defined and open ended problems requiring divergent thinking
- synthesis of material outside normal boundaries
- evaluation – technical decisions vs. social and ethical considerations
- problem finding and definition (not just problem solving)

This model is intrinsic to the PDE educational approach where creativity and innovation are emphasised. Engineering education that focuses on problem solving results in graduates who can resolve a technical issue, but who are not solution focussed. Product design innovation does not result from solving a problem; it emerges from a creative, explorative and reflective process that expects only the best, a unique and creative solution.

Sketching

The product design engineering pedagogy through its integration of engineering with 'designerly ways' ensures that these new engineers are fluent at creative exploration and critical reflection and can 'handle different levels of abstraction simultaneously." (Cross, 2000) This is achieved through the requirement for aptitude in perspective sketching, which is taught from the first semester, and developed continuously throughout the four year undergraduate program. Students are required to utilise sketching in all stages of a design proj-

Figure 2. Innovative product solutions (Device to apply chemically-impregnated tape to grapevines to eradicate environmental contamination (left) Shared ownership urban 2020 vehicle concept utilising a sustainable propulsion system (right))



ect; beginning with initial conceptualisation, then concept development (Figure 3) and progressing to detail design and technical resolution (Figure 4).

In the PDE curriculum, perspective sketching and rendering are integrated into the learning process at all levels of the course experience and students are expected to become fluent communicators of form and function through non-digital drawing media.

Socially Responsibility

To address these goals, the Swinburne product design engineering program has established several strategic relationships with humanitarian aid agencies with the view to providing proficient design and engineering services in areas of desperate need, whilst providing students an opportunity for real world learning and experience in design for good. Product design projects have been conducted addressing scenarios provided by World Vision (a leading humanitarian aid agency) (Figure 6) and liaison with Engineers without Borders has been established to engage in long term infrastructure building activities.

Final year 'capstone' projects are required to address a social need through a product outcome. Students are encouraged to collaborate closely with partners from aid agencies, medical research organisations, disability support organisations and education and healthcare specialists to provide product solutions to real world needs (Figure 5). Such projects have developed an ethical and responsible approach to the practice of engineer-

Figure 3. Socially responsible designs. (Mobility aid/shopping trolley for elderly users (left), Ceramic pressurized sterilization unit for Nepal designed for local production (right))

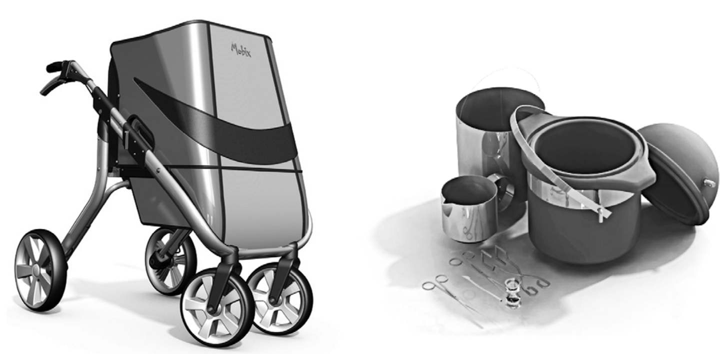


Figure 4. 2nd year project work – sustainable electronic products. (Solar-powered soil analysis device (left); Hand-powered communication/ banking access device for remote communities (right))

ing design amongst the student cohort, and many altruistic endeavours post-graduation.

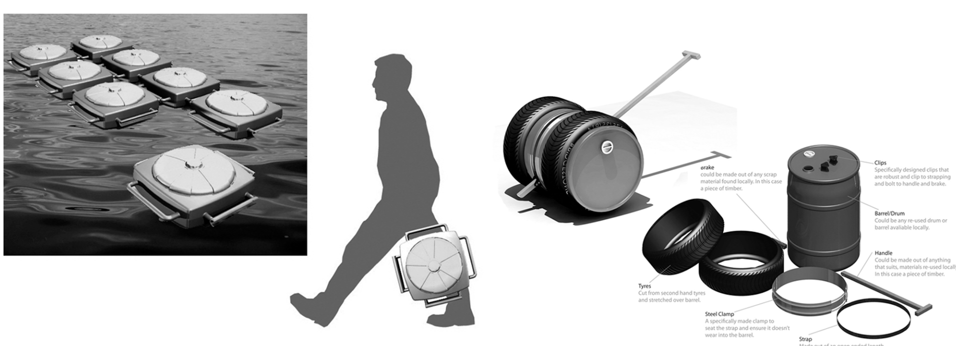
Sustainability

The product design engineering program focuses on sustainable design from the second year where the students are introduced to the principles of sustainability, the importance of life cycle analysis (cradle to grave), and ethical and eco-design methods. Students examine the impact of e-waste and the WEEE directive and then utilise environmental design tools to develop environmentally sustainable electronic devices (Figures 7 and 8). Eco-design is integrated from the early conceptual stages to ensure appropriate and well considered designs; band-aid solutions and products without a 'sustainable purpose' or genuine need are not accepted. Design outcomes must adhere to the "Ten Golden Rules of Eco-Design" (Luttropp & Lagerstedt, 2006), life cycle analysis (using Eco-Indicator and other software) and 'end of life' scenarios must be considered.

Open Ended Problems

It is important that students are set open–ended challenges where they must address a real world scenario without clear definition of potential outcomes. A typical project might be the design of a water transportation device to enable remote communities to convey water across either rug-

Figure 5. Open-ended problem solving exercise: water transportation for remote communities (Modular water transportation system for use in flooded landscapes (left); Water transportation system utilising readily available discarded materials (right))



ged or flooded terrain (Figures 9 and 10). These 'wicked' problems have numerous potential outcomes and are not constrained by known entities such as materials or manufacturing processes. Instead the students must consider the user and environment, define the problem and then design the product solution.

These projects are invaluable tools to develop creativity, and as students become familiar with uncertainty and confident in their design ability, true innovation emerges.

Future Trends in PDD: A Changing Environment

Industry trends indicate that product design and development will increasingly require an interdisciplinary professional capable of operating with distinction in both design and engineering roles. It is already apparent that successful design results from teams that integrate the skills of both engineering and industrial designers through non-combative collaboration built on mutual respect.

The challenges of the next few decades are immense; design teams must lead behavioural change through well considered, appropriate and sustainable products that meet the future needs of all stakeholders, not just their clients and customers. Confronting climate change inflicted by environmental degradation, reducing the impact of products throughout their life cycles, and enhancing the health, life expectancy and aspirations of those at the base of the pyramid will require real innovation in product design and development. This innovation will not occur naturally, a paradigm shift is required; one that will challenge expected conventions in PDD and revolutionise product design and manufacturing activities.

Advances in solid freeform fabrication (SFF) or rapid manufacturing (RM) have the potential to dramatically reshape the existing manufacturing sector. New design protocols and working methodologies will be developed to maximise the potential of these emerging manufacturing processes. The advent of rapid manufacturing will profoundly impact existing global manufacturing systems and may precipitate a move to local, or even customer production, using SFF technologies. As components and products are freed from the constraints of traditional manufacturing processes, fresh and innovative product architecture will emerge: we are at the dawn of a new product vernacular. Without the restrictions of centralised manufacture, transport and storage considerations will no longer critical as localised 'just in time' or 'buy and make' production of 'end use parts' will drive new directions in product design.

The implications of RM for the developing world are immense; previously unheralded access to products and services will drive rapid societal change. The removal of transportation from the product supply chain would dramatically reduce costs and aid in greenhouse gas reduction and fossil fuel consumption. Manufacturing processes not reliant on immense tooling and infrastructure costs will allow product design teams' far greater freedom to explore product potential and even to customise products for individual markets and/or customers. With this emerging technology, the traditional shackles of design for manufacturing will be removed or at least will be much less constraining, and this will revolutionise the product design and development environment.

The existing system of product manufacture, which has its origins in the Industrial Revolution, will be rapidly transformed and these changes may have similar socio-economic consequences and cultural impact, to that of mechanisation in the late 18th century.

This new environment will see the emergence of a new creative PDD professional; adaptable flexible, multilingual in the languages of engineering and design, responsible and culturally sensitive. The product design engineering paradigm which integrates engineering and design to develop an interdisciplinary professional, aims to produce graduates for these challenging and uncertain times; design engineers who are well equipped to adapt quickly to shifting environments and new methodology, in the pursuit of socially responsible and sustainable product innovation.

CONCLUSION

This chapter has dealt with the challenges facing product design and development teams in the near and immediate future. PDD needs to respond to existing and emerging challenges with a new paradigm that incorporates fresh approaches to practice that respond to global societal and environmental needs, empower individuals and cultures and embrace emerging technologies that will impact significantly on established protocols and the definition of professional disciplines.

The role of engineers in the product design and development process is shifting and will continue to be redefined as technological, environmental and societal considerations impact on professional practice. It is felt that these challenges are best addressed with a complex and multi-skilled practitioner, who bridges traditionally rival professions and facilitates a new era of collaboration, understanding and respect. Both the engineering and design professions need redefinition, a paradigm shift to address the rapidly changing climate with which we are faced. Engineers have always been integral to the PDD process, but their roles are shifting and the importance of their contribution in the near future, should not be underestimated.

The product design and development process can be enhanced by innovative engineers, but often is hampered by inflexibility, poor understanding of user needs and cultural insensitivity. The PDD environment will be the beneficiary of a more enlightened (and less rigid) approach to engineering education that seeks to impart skills that result in responsible practice and creative outcomes. And both society and environment will benefit from an engineer who is creative, innovative, adaptable and responsible; who can contribute to appropriate product design through new professional synergies and guide attitudinal change.

The role of engineering education in appropriate product design and development cannot be underestimated. Educators have a responsibility to global communities to produce engineering graduates with embedded ethical and creative philosophies who are well prepared and eager to make a significant contribution to environmental and societal issues. Early indications are that the product design engineering paradigm (emerging in direct response to the needs of industry) contributes to future global needs through a new engineering pedagogy that imbues its students with a strong understanding of the importance of their future contributions to society and the impact of appropriate professional behaviour.

The world is expectant of product design and development to address the future needs of our societies and environments. This will require new skills, enhanced communications and new synergies between stakeholders. The next generation of engineers, properly prepared, can direct the product design and development process confidently in new strategic directions.

REFERENCES

Akay, A. (2003). The renaissance engineer: educating engineers in a post-9/11 world. *European Journal of Engineering Education*, *28*(2), 145–150. doi:10.1080/0304379031000078979

Brunel University. (2009). *Product Design Engineering - Course Summary*. Retrieved in 2009, from www.brunel.ac.uk/courses/ug/cdata/p/ProductDesignEngineeringBSc

Cropley, D. H., & Cropley, A. J. (2000). Fostering Creativity in Engineering Undergraduates. *High Ability Studies*, *11*(2), 207–218. doi:10.1080/13598130020001223

Cross, N. (2000). *Engineering Design Methods: Strategies for Product Design* (3rd ed.). New York: Wiley.

Cross, N. (2001). Designerly ways of knowing: design discipline versus design Science. *Design Issues*, *17*(3), 49–55. doi:10.1162/074793601750357196

Design Victoria. (2008). *Five Years On: Victoria's Design Sector 2003-2008*. Melbourne: Retrieved from http://www.designvic.com/~/media/PDF/FiveYearsOn.ashx

Dym, C. A. (1999). Learning Engineering: Design, Languages and Experiences. *Journal of Engineering Education*, *88*(2), 145–148.

Dym, C. A., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering Design Thinking, Teaching, and Learning. *Journal of Engineering Education, January 2005*, 103-119.

Eekels, J. (1987). Guidelines for Engineering Teachers Concerning Educating the Engineer for Innovation and Entrepreneurial Activity. *European Journal of Engineering Education*, *12*(3), 259–270. doi:10.1080/03043798708939369

Eekels, J. (1994). The Engineer as Designer and as a Morally Responsible Individual. *Journal of Engineering Design*, *5*(1), 7–23. doi:10.1080/09544829408907868

Felder, R. M. (1987). On Creating Creative Engineers. *English Education*, *77*(4), 222–227.

Fox, R. (1981). Design based studies: an action-based 'form of knowledge' for thinking, reasoning and operating. *Design Studies*, *2*(1), 33–40. doi:10.1016/0142-694X(81)90027-2

Fry, R. (2006). *Defining the Obvious: Explaining Creativity and Design Thinking to Nondesigners*. Paper presented at the 2006 IDSA National Education Conference.

Ghosh, S. (1993). An Exercise in Inducing Creativity in Undergraduate Engineering Students Through Challenging Examinations and Open-Ended Design Problems. *IEEE Transactions on Education*, *36*(1), 113–119. doi:10.1109/13.204828

Glasgow School of Art. (2009). PDE Program - Relevance to Industry and Society. Retrieved from www.gsa.ac.uk/gsa.cfm?pid=1421

Goel, V. (1995). *Sketches of Thought*. Cambridge, MA: MIT Press.

Goldschmidt, G. (1991). The Dialectics of Sketching. *Creativity Research Journal*, *4*(2), 123–143. doi:10.1080/10400419109534381

Grasso, D., & Martinelli, D. (2007). Holistic Engineering. *The Chronical Review*, *53*(28), B8.

Green, G., & Kennedy, P. (2001). Redefining Engineering Education: The Reflective Practice of Product Design Engineering. *International Journal of Engineering Education*, *17*(1), 3–9.

Hammer, J. (2007). Sustainability - an Introduction from the National President. Retrieved November 12, 2008, from http://www.engineersaustralia.org.au/ieaust//quicklinks/sustainability.cfm

IME. (1985). *The Formation of Mechanical Engineers: Present and Future Needs* (The Grant Report).

Lipson, H., & Shpitalni, M. (2000). Conceptual design and analysis by sketching. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *14*, 391–401. doi:10.1017/S0890060400145044

Luttropp, C., & Lagerstedt, J. (2006). EcoDesign and The Ten Golden Rules: generic advice for merging environmental aspects into product development. *Journal of Cleaner Production*, *14*, 1396–1408. doi:10.1016/j.jclepro.2005.11.022

Østfold University College. (2009). Study Program for the Engineering Education: Product Design Engineering - course description. Retrieved from www.hiof.no/index.php?ID=1760&lang=eng

Papanek, V. (1985). *Design for the Real World: Human Ecology and Social Change* (2nd ed.). Chicago: Academy Publishers.

Pappas, E. (2002). *Creative Problem Solving in Engineering Design*. Paper presented at the ASEE Southeast Section Conference

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. doi:10.1007/BF01405730

Santamarina, J. C. (2002, February 21-22, 2003). *Creativity and Engineering - Education Strategies.* Paper presented at the International Conference/Workshop on Engineering Education Honouring Professor James T. P. Yao, College Station, Texas.

Stanford University. (2009). *Product Design - course description*. Retreived from http://design.stanford.edu/PD/undergraduates.html

Stouffer, W. B., Russell, J. S., & Oliva, M. G. (2004). *Making the Strange Familiar: Creativity and the Future of Engineering Education*. Paper presented at the 2004 American Society for Engineering Education Annual Conference and Exposition.

Swinburne University of Technology. (2009). *Product Design Engineering - course description*. Melbourne, Australia: Swinburne University of Technology.

Ullman, D. G., Wood, S., & Craig, D. (1990). The Importance of Drawing in the Mechanical Design process. *Computing and Graphics*, *14*(2), 263–274. doi:10.1016/0097-8493(90)90037-X

Ulrich, K. T., & Eppinger, S. D. (2004). *Product Design and Development* (3rd ed.). New York: McGraw-Hill/Irwin.

United Nations, W. C. o. E. a. D. (1987). Our Common Future. *Report of the World Commission on Environment and Development, General Assembly Document A/42/427*

University of Strathclyde. (2009). *Product Design Engineering-course description 2009*. Retrieved from www.strath.ac.uk/dmem/studying/undergraduatecourses/choosingtherightcourseforyou/

Verstijnen, I. M., & Hennessey, J. M. (1998). Sketching and Creative Discovery. *Design Studies*, *19*(4), 519–546. doi:10.1016/S0142-694X(98)00017-9

KEY TERMS AND DEFINITIONS

Base of the Pyramid (BoP) - the other 90%: The base of the pyramid refers to the largest but poorest socio-economic groups; the four billion people who survive on less than $2 per day. These groups are not only isolated from the affluent materialistic lifestyles of first world Western populations, but are adversely affected by our development, manufacturing and quest for diminishing resources, and are the earliest victims of environmental degradation. Central to assisting those at the BoP are the principles of understanding, empowerment and respect, participatory development, mutual value and co-creation. Many of 'the other 90%' lack access to clean drinking water (1.1 billion), adequate sanitation (1.4 billion), healthcare and education. There exists many opportunities for design to contribute to improving the life expectancy and aspirations of those in need, through socially responsible design.

Creativity: A process that involves generation of new ideas and concepts; resulting in unique, innovative approaches and solutions. Creativity is an essential element of design thinking and depends on confidence to explore the unfamiliar, tolerate the unexpected and encourage the unknown. Creativity is an expressive and personal process that involves divergent thinking, exploration and reflection and at times relies on intuitive action.

Open-Ended Problems: Facilitate the development of creative problem solving skills. They are typically ill-defined problems that do not have an obvious single correct answer, and require a unique solution that emerges from a creative process. Open ended problems often require intuitive responses from students and the acceptance of the unexpected, which can cause discomfort for those with underdeveloped critical thinking or creative ability. It is critical that designers and engineers experience many of these 'wicked' problems throughout their education in order to develop problem framing skills, confidence in divergent thinking, and lose the tendency to fixate on prior solutions.

Pedagogy: Refers to strategies of teaching or principles and methods of instruction, sometimes referred to as 'the art of teaching'. It is effectively an array of teaching strategies that support the learning experience, through a well considered approach to curriculum development that empowers the teacher, student and learning experience. In this chapter, pedagogy is used as a broad descriptive term when describing the teaching strategies and curricula of the product design engineering paradigm.

Product Design Engineering (PDE): Product design engineering is an educational pedagogy and professional engineering paradigm, which seeks to integrate two disparate and traditionally rival professions; mechanical engineering and industrial design. Product design engineers are a new breed' of design engineer, interdisciplinary, creative and human centred, and highly skilled in the product design and development arena.

Project Based Learning (PjBL): An experiential learning process that involves the use of classroom projects to stimulate deep learning through enquiry and experimentation. In the context of design education it involves the students undertaking a design process in response to a detailed brief usually in a way that replicates real world professional practice. Project bases learning requires students to engage in design activities including conceptualisation, reflection, problem solving and exploration of possible solutions and detail resolution. The learning process is defined partly by the final artefact (design) and the success of the methods applied. In the context of teaching engineering design, PjBL allows students to integrate scientific knowledge into a solution-focussed approach encouraging 'learning through doing' which entails increased knowledge retention, and facilitates development of skills in the 'practice' of engineering' rather than relying on memorisation of theory.

Rapid Manufacture (RM): Rapid Manufacture utilises processes originally known as rapid prototyping (RP) and employs solid free-form fabrication (SFF) to deliver three dimensional components. It is an additive fabrication method that utilises sequential delivery of material and/or energy to specified areas to construct a particular form, layer by layer. As this technology evolves, increasingly complex forms are possible, including shapes and assemblies not achievable through traditional fabrication processes. This 'next level' technology has the potential to revolutionise manufacturing with particular regard to localised or even customer manufacture through advanced '3D-printing' processes.

Socially Responsible Design (SRD): The use of design to address social, environmental, economic issues and focuses on a move to move beyond first world consumer demands towards a more holistic and responsible approach to product design that embraces ethical, cultural and humanitarian values; simply termed 'design for good' or 'design for need'. SRD responds to all stakeholders (not just clients and customers) and examines the consequences of design activity and the potential for design to contribute positively to societal aspirations and expectations, health and lifestyle.

Sustainable Design: Sustainable design addresses the 'triple bottom line', addressing the principles of economic, social and environmental sustainability. It aims to counter negative impacts of products throughout their lifecycle with 'cradle-to-grave' philosophy that employ low impact materials and processes, and considers energy usage, embedded energy, design for disassembly to facilitate reuse or recycling, resource renewability, bio-mimicry and design impact measures (such as life cycle analysis). Sustainable design aims to support development "that meets the needs of the present without compromising the ability of future generations to meet their own needs." (United Nations, 1987)

This work was previously published in Handbook of Research on Trends in Product Design and Development: Technological and Organizational Perspectives, edited by Arlindo Silva and Ricardo Simoes, pp. 173-194, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 6
E-Learning for SMEs:
Challenges, Potential and Impact

Asbjorn Rolstadas
Norwegian University of Science and Technology, Norway

Bjorn Andersen
Norwegian University of Science and Technology, Norway

Manuel Fradinho
Cyntelix, the Netherlands

ABSTRACT

SMEs have a special need to be able to collect knowhow from the global community and implement this as competence in the enterprise. E-Learning is an approach for competence development that can assist SMEs in creating a learning organization. There are different approaches to E-Learning. Amongst the most powerful ones is the application of serious games. The PRIME project conducted experiments with serious games in project business and in manufacturing. This has resulted in a set of guidelines for successful implementation of games in an organization. It is necessary to allow a learning curve and put emphasis on the social context to obtain effective learning. For delivery of E-Learning the hybrid learning model is recommended. SMEs are advised to develop E-Learning in cooperation with academic institutions, and to use real life cases and problems for assignments. Delivery should be based on proven technology when used in SMEs.

INTRODUCTION

The extended enterprise is being implemented worldwide using different e-business technology. To be on the competitive arena today requires cutting-edge competence not only on the products supplied, but also in all aspects of business operation and competence development. ICT is in this respect an extremely powerful tool enabling enterprises to work globally and allowing creation of supply chains, enterprise networks, and learning communities. For SMEs this represents a challenge. SMEs have limited resources to drive this technology, but are at the same time dependent on access to the same technology. There is a special need to be able to collect knowhow from the global community and implement this as competence in the enterprise.

DOI: 10.4018/978-1-4666-1945-6.ch006

E-Learning is an approach for competence development that can assist SMEs in creating a learning organization. E-Learning allows learning at the time, place, and pace decided by the learner. It enables tailoring of learning content to the needs of the user and facilitates collaborative learning and access to expert teachers wherever they might be. Recent research from previous and on-going EU-projects (GEM, PRIME, and TARGET) has been able to demonstrate some of the power of E-Learning, or more correctly "technology-enhanced learning". Some key research findings from these projects will be presented together with a case study on practical implementations of E-Learning solutions in industry.

The GEM project developed a framework of a curriculum for manufacturing strategy. The project included a study of different delivery methods. The PRIME project used serious games to train middle management in decision making. It was tested through demonstrators for repetitive manufacturing and for project based business. The TARGET project is ongoing and is studying different learning approaches based on concepts for technology-enhanced learning (TEL).

The objectives of this chapter is to discuss different E-Learning approaches that are relevant for SMEs based on research findings from the mentioned projects and to develop an agenda for future research for competence development in SMEs using advanced ICT. It indirectly explore how SMEs can exploit E-Business by discussing how innovative E-Learning approaches can be used to facilitate transfer of knowledge and building of experience. This new knowledge and experience can in turn be leveraged by SMEs to develop and experiment with E-Business strategies.

E-LEARNING BACKGROUND

There exists a number of questionnaire-based evaluation approaches for E-Learning (Sage, 2002, U.S.I.G., 2003, Carliner, 2002, Hughes, 2003). They indicate four essential areas for evaluation (Schwesig, Rolstadås, and Thoben, 2005):

- E-Learning content - the preparation and the selection of content
- E-Learning approach - the instructional approach of the E-Learning course
- E-Learning delivery mechanism - the actual design of the learning environment
- Overall assessment- general questions, strengths and weaknesses, suggestions for improvement for the course

In this context, we will focus on E-Learning approach and delivery mechanisms.

E-Learning Approaches

Most SMEs are, in terms of organizational structure, fairly simply designed, with few departments or units. On the other hand, this means that typically employees are involved in more than one function or discipline required to run the company. In such integrated areas, decisions are taken that have a crucial influence on the competitiveness and profitability of a company. These decisions relate to questions like various strategic management tasks, product portfolio, product program, competitor strength, competition parameters, market segmentation, project planning, innovation management, process management, etc.

These integrated areas all require competences that go beyond the narrow technical competences that can be acquired easily and affordably through traditional courses or training. As complexity and dynamics in the internal and external specialization and collaborations increase, the need for training and competence development increases as well. In terms of competence development, organizations therefore have concrete needs that are not met well by traditional learning approaches:

- Reduce the time it takes an employee to acquire the necessary competences to do

their job in the most efficient and effective manner;
- Change the learning context rapidly and in response to the real world;
- Facilitate knowledge sharing within an organization;
- Support a soft failure environment where mistakes have no impact on the real world, thus promoting a willingness to engage in measured risk taking, focused on achieving a high level of polished performance in the real world

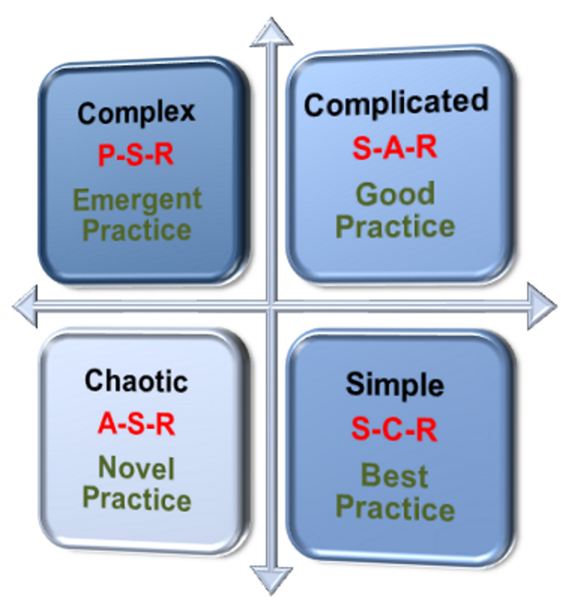
The above are a response to market pressures, but society itself has also gone through changes as a result of the advances of technology. For example, the way people scan reading material has changed significantly due to the World Wide Web, thus having a significant impact on marketing strategies (Meyer, Chall, Onofrey, and Rose, 1998). With E-Learning it is no different and already there are signs that E-Learning needs to take account of the enhanced abilities of the so-called gamer generation, such as differences in cognitive processes when compared to non-gamers, which can give gamers a competitive edge in fast-changing market situations such as arise in economic downturns (Herring and Job, 2008). This has implications for organizations that need to change work environments and business practices to face new challenges. In this light games can be seen as the most recent element in E-Learning offering new and different approaches to organizations in their effort to upgrade and enhance competences.

Traditionally, problems have been seen as complicated challenges that should be solved through breaking them down into smaller and smaller chunks. Seen from the solution perspective, the solutions then emerge by solving the problems related to the smaller chunks. Largely, this reflects an analytical way of thinking. In such regimes, learning and training are also related to the smaller chunks. By training the competences behind the smaller chunks the person gradually becomes able to meet the challenges.

We experience that most modern problems are more frequently complex rather than complicated. Complex problems are messier and more ambiguous in nature; they are more connected to other and often very different problems; more likely to react in unpredictable non-linear ways; and more likely to produce unintended consequences. Most organizations have been designed to deal with a complicated rather than a complex world. Hierarchical and silo structures are perfectly designed to break problems down into more manageable fragments. They are not, however, so effective in handling high levels of complexity. For this reason many institutions and companies are now struggling to adapt to a more complex world.

The Cynefin (the word has a Welsh origin and is pronounced *kun-ev'rin*) framework proposes an association between the nature of context of problems and appropriate responses (Snowdon & Boone, 2007). The Cynefin framework consists of five domains (see Figure 1 for an adapted version):

Figure 1. The Cynefin Framework



- Simple, in which the relationship between cause and effect is obvious and the appropriate approach is to Sense – Categorize – Respond
- Complicated, in which the relationship between cause and effect requires analysis or some other form of investigation and/or the application of expert knowledge, and the appropriate approach is Sense – Analyze – Respond
- Complex, in which the relationship between cause and effect can only be perceived in retrospect, but not in advance, and the appropriate approach is Probe – Sense – Respond
- Chaotic, in which there is no relationship between cause and effect at systems level, and the appropriate approach is to Act – Sense – Respond
- Disorder – in which it is unclear what type of causality exists. Disorder is in the centre of the framework. Here people will revert to their own comfort zone and choose the approach related hereto.

Each domain of the Cynefin framework represents different levels of expected achieved practice. In the simple domain, we can expect "Best Practice". In the complicated domain, we can expect "Good Practice". In the complex domain, we can expect "Emergent Practice", and in the chaotic domain, we can expect "Novel Practice".

These expectations relate to the nature of the problems. Heavily process-oriented situations, such as loan payment processing, are often simple contexts. Directives are straightforward, decisions can be easily delegated, and functions can be automated. As problems become complicated, they tend to require assistance from specialists: A car owner may know that something is wrong with his car because the engine is knocking, but he has to take it to a mechanic to diagnose the problem.

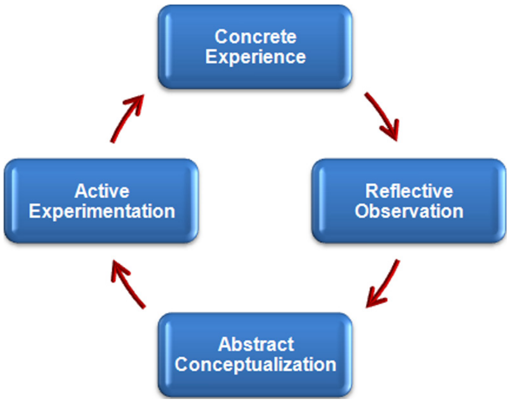
When problems become complex, the clear pattern between cause and effect disappears. We can understand why things happen only in retrospect. Instructive patterns, however, can emerge if the leader conducts experiments that are safe to fail. That is why, instead of attempting to impose a course of action, we must patiently allow the path forward to reveal itself. We need to probe first, then sense, and then respond. There is a scene in the film Apollo 13 where the astronauts encounter a crisis that moves the situation into a complex domain. A group of experts is put in a room with a mishmash of materials – bits of plastic and odds and ends that mirror the resources available to the astronauts in the flight. The team is told: This is what you have; find a solution or the astronauts will die. None of those experts knew a priori what would work. Instead, they had to let a solution emerge from the materials at hand.

In the chaotic context, searching for right answers would be pointless: The relationships between cause and effect are impossible to determine because they shift constantly, and no manageable patterns exist – only turbulence. Here we must first act to establish order, then sense where stability is present and from where it is absent, and then respond by working to transform the situation from chaos to complexity, where the identification of emerging patterns can both help prevent future crises and discern new opportunities.

The analytical learning and competence building approach that has proven efficient when dealing with simple and complicated problems proves inefficient when confronted with complex problems. In short the "probe" element is missing in traditional approaches, and the "sense" element appears to be very different when dealing with complex problems compared to complicated and simple problems.

The probe and sense elements seem to be crucial when dealing with complex problems. This is supported by a number of research results published in the past 30 years. In 1984 David Kolb published his book title "Experimental Learning" (Kolb, 1984). His claim is that we are learning by cyclic patterns of four types of activities: Concrete

Figure 2. David Kolb's learning circle



experience, reflective observation, abstract conceptualization, and active experimentation (see Figure 2). According to Kolb it does not matter where we start - the important requirement for real learning is that one goes through the full cycle.

Though Kolb's work is aiming at the individual level there are several contributions supporting that a similar pattern is valid at the organizational level. Dorothy Leonard argues that: "The primary activities spawning organizational learning are experimentation and prototyping" (Leonard-Barton, 1995) and Argyris and Schön (1978) have introduced the notion of single-loop and double-loop learning which includes active experimentation.

Donald Schön has been studying how professionals are working very differently from novices. His point is that when people have reached a certain level of professionalism it will change their working style and they become "reflecting practitioners" (Schön, 1983). The reflective practitioner is in a constant process of thinking, reflecting, acting, and building experience – very much in line with the learning process as described by Kolb. This process is efficient for the professional person but due to the amount of tacit knowledge it is often difficult to articulate and share the results with others (Polanyi, 1967). Physical models or other model representations seem to be the most efficient means to facilitate this sharing (and learning) process (Schrage, 2000).

In his book, Serious Play, Michael Schrage (2000) praises many aspects of prototyping for speeding up processes etc. and mentions examples of great breakthroughs made by new prototyping tools. The following extracts provide exemplary viewpoints from the book:

- When talented musicians improvise, you don't look inside their minds; you listen to what they play. When talented innovators innovate, you don't listen to the specs they quote. You look at the models they've created.
- The challenge of converting uncertainty into manageable risks or opportunities explains why serious play is often the most rational behavior for innovators.
- Serious play is about improvising with the unanticipated in ways that create new value.
- Prototypes engage the organization's thinking in the explicit. They externalize thought and spark conversation.
- Prototypes force confrontation with the tyranny of trade-offs.
- The conventional wisdom that "innovation processes" drive prototype development is misleading. Empirical observations of organizations with effective innovation cultures confirm just the opposite: changes in prototypes and simulations drive the innovation process.
- Prototypes are machine tools for producing choice.
- Most companies have formal prototyping processes and informal prototyping cultures.

Schrage (2000) argues against the common assumption that "great teams make prototypes" and suggests that instead one should realize that "prototypes make great teams". The making of great teams goes beyond the individual team, but helps create teams out of people with different backgrounds by creating "shared space". Shared space is the common ground where people can meet on even terms and objectively discuss matters.

Obviously, physical and mental prototypes provide opportunities to probe and sense. They invite for questioning and challenging of conventional wisdom (March, 2006) and thereby learning and competence development can be supported.

One of the most promising, future oriented concepts of E-Learning is the use of serious games. (This concept will be described more detailed in a later section.) The game culture has been studied by Oliveira, Andersen, Oliveira, and Rolstadås (2006). They assessed the culture and its impact using a questionnaire consisting of 48 questions. The results of the survey demonstrate that the gamer generation is not as prominent in Europe as in the United States, but present nonetheless. The managers, who experienced an upbringing with the influence of digital games, demonstrated an advantage over those that did not share the same background. Throughout all the statements aimed at capturing a respondent's thoughts and motivation on particular topics, the gamers revealed an edge and advantage over those without gaming experience.

Delivery Mechanisms

Delivery of E-Learning relies heavily on Internet-based technologies (Rolstadås, 2002; Lefrere, 2007). Technology, the fast outdating of knowledge and training, the desire for just-in-time training delivery, and the continuous hunt for cost-effective methods to meet learning requirements have redefined the processes of design, development, and delivery of training and education (Urdan and Weggen, 2000). Several terms have been used to characterize the new methods of E-Learning including: technology-enhanced learning, distributed learning, online learning, and web-based learning. Many educational organizations are experimenting with E-Learning, using intranets, websites, and multimedia solutions, and a great deal has been written about the issues and problems encountered in these attempts (Hall & Snider, 2000; Urdan & Weggen, 2000; Berge, 1998). However, by giving careful consideration to the design and implementation of E-Learning, successful courses that take full advantage of the benefits enabled by the technology can be created.

Delivery of E-Learning content can be synchronous or asynchronous. The alternate use of video, audio, pictures, text, (interactive) animations, quizzes and chat, and discussion boards should assure that both styles can be covered (Rolstadås and Hussein, 2002). Whatever the choice of learning communication, the learning success heavily depends on the actual form of mediation. According to Dale (1969), passive learning activities like reading texts or watching a demonstration result in the fact that just 10% (reading) or 30% (watching) of the content is remembered afterwards. Active involvement within the learning process, like participating in a discussion, leads to the fact that the learners can remember about 70% of the related content afterwards (Dale, 1969). Accordingly, content needs to be delivered on the basis of active learning. In order to assure the appropriate employment of the mentioned design elements and the application of active learning elements, design guidelines have been used explaining the correct use of pictures and texts, (interactive) animations, quizzes, chat/messaging video, and audio. Kommers, Grabinger, and Dunlap (1996) present a number of design elements such as pictures and text, animations, quizzes, chats, messaging, video, audio, games, and laboratories that can bring maximum benefit to the learner.

A powerful concept for delivery is based on the concept of the hybrid learning published by Rolstadås and Hussein (2002). They claim that this type of mixed instructional delivery has proved to be an effective tool for enhancing the learning value for the workforce. Applying this approach may also lead to significant reduction in learning time. The hybrid solution is well suited for large organizations distributed over a large geographic area. Thus, the target group may be characterized by diverse social, cultural, and academic background. Therefore, it is important that the model enables multiple learning styles.

According to Rolstadås and Hussein (2002), the hybrid model must contain the following properties:

- Create an environment for learning that addresses multiple learning styles
- Ensure active participation and to focus on accommodating the learners' former experiences
- Enable the learners to build personal and possibly social networks
- Use information and communication technology for designing the content in order to enrich the learning experience
- Restrict the use of on-campus delivery whenever high level of interactivity between the learners themselves is required or between the learners and the content
- Availability in the form of an updated support system that assists the learners on the academic, technical, and administrative level
- Learners should be motivated and self-disciplined and using education in order to achieve clear goals.

The hybrid model is based on a mixture of on-campus plenary learning sessions (face to face), self paced learning using rich media through virtual classroom, and collaborative group work using collaborative ICT tools included in the virtual classroom.

IMPROVING BUSINESS DECISION CAPABILITY USING SERIOUS GAMES

The knowledge-based economy has redefined the rules of economic development with continual and ever-more rapid industrial and social changes. Enterprises have realized that knowledge workers are a key to a company's productivity, competitive edge, and ability to adapt to changes in markets and customers. A highly educated, well-trained, and adaptable workforce is a key element in competitiveness, productivity, and the growth of employment, thus life-long learning (Aspin and Chapman, 2000) anywhere and anytime has become a fundamental strategy of the knowledge-based organisations. This is corroborated by the report of the Task Force on Employment, chaired by Wim Kok (2004), and the Spring European Council of 2004 stressed that Europe must invest more, and more effectively, in its human capital. These recommendations are clearly guiding the revised Lisbon Knowledge Society initiative, the i2010 Strategy. However, tailoring effective E-Learning frameworks within organizations is highly challenging and its promises have not been highly successful in addressing the challenge of supporting life-long learning of knowledge workers. Within SMEs, the challenges of deploying successfully an E-Learning framework are exacerbated due to their focus on immediate business survival.

As pointed out by Romiszowski (2004), the failure of E-Learning can be associated to some common mistakes, which are rooted in the focus on the technology itself rather than on the impact of technology in the context of human-activity systems (i.e., education and training systems). A promising technological innovation in E-Learning platforms is the use of digital games, which according to Prensky (2001) introduces play into the learning process, an essential missing factor in traditional technology-enhanced learning (TEL) platforms since learning is more than education.

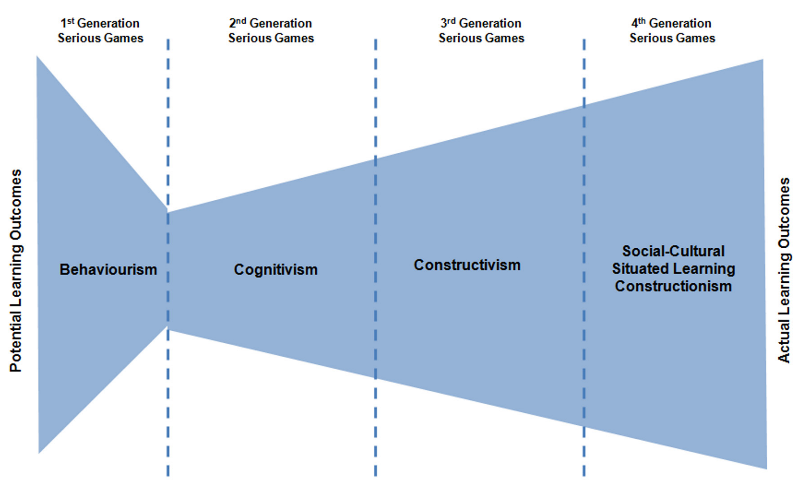
In fact, the most common term currently in use is serious games, where the flow engagement of digital games is merged with concrete learning outcomes. Although only recently have serious games gained more prominence in mainstream learning environments, one may consider that a total of 3 generations have taken place with a 4th generation emerging (Figure 3).

The first generation focused on learners acquiring designated behaviors through conditioning and reinforcements. The approach is only interested in directly observable behavior based on stimuli and response. The learning theory does not take into account that learners are different from one another and assumes that a learner acquires a skill provided it is practiced enough times. This led to the second generation where the focus shifts from the behavior to the learner and their internal cognitive constructs to represent knowledge. However, the feasibility of knowing precisely the learner's cognitive processes is limited. The third generation relies on constructivism, which still focuses on the learner, but from the outside. In this case, the learners build and constructs their knowledge by interacting with the environment and each learner acquires the knowledge in their own way. Finally, the current generation of serious games is based on situated learning where constructivism is complemented by social processes that are necessary for learners to externalize knowledge.

PRIME Case Description

Providing Real Integration in a Multi-disciplinary Environment (PRIME) is a serious game on the borderline between the 3rd and 4th generation, but exploiting some of the key strengths of cognitivism. The PRIME serious game was the result of a European IST project (FP6-016542), that aims to give business professionals in strategic manufacturing a learning environment where they can experiment with new ideas and learn the effects of their decision making. The PRIME serious game consists of a Virtual Business Environment (VBE) that defines the global business reality where users are responsible for the management of a single Business Unit (enterprise). A crucial characteristic, if not the strongest contributing factor, to learning is the feedback capability available to the learner. In addition, emotions (fun) are also connected to learning, where "people learn in direct proportion to the amount of fun they are having" (Rose and Nicholl, 1997). The

Figure 3. The four generations of serious games according to their driving learning theories

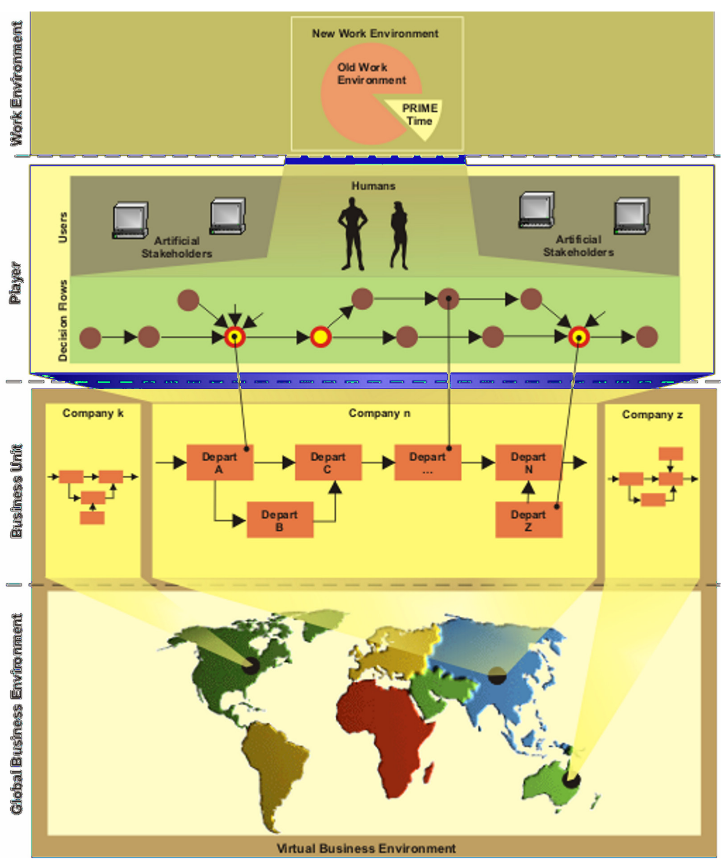


VBE provides real-time feedback to all the user's interactions and being a serious game, it focuses on providing a fun and engaging environment for users. The management breakdown of a business unit is based on eight functional units: Production, Product Development, Sales, Human Resources, Strategic Marketing, Distribution, Finance and Information Systems. Associated to each functional unit is a set of operation processes that are based on an input-transformation-output model.

The block diagram of Figure 4 corresponds to the PRIME conceptual model, illustrating the interplay of the global business environment and the different business units, each controlled by either a human player or an artificial stakeholder (agent).

The heart of the PRIME Concept Model is the set of Decision Flows where each flow consists of a linked graph of strategic decisions, which have far-reaching consequences for an organization and typically involve the commitment of vast resources. The decisions play out over long time frames and have significant opportunity costs. Strategic decisions often must be made under conditions of substantial uncertainty, particularly when complex policy objectives must be reformulated in the face of a dynamic, sometimes volatile strategic environment. Initial assumptions about the environment and other players may be incorrect or incomplete. The Decision Flows were shaped by applying a methodology based on Critical Incident Technique (CIT) (Flanagan, 1954) to capture the decisions supported by the PRIME Serious Game, which can be aggregated together in the following main categories: Logistics, Product Innovation and R&D, Sales and

Figure 4. PRIME Conceptual Model



Strategic Marketing, Production, and Finance. In essence, the Decision Flows define the game play of the PRIME serious game, which is an innovative approach to defining a serious game.

The implementation of the VBE relies on a hierarchical simulation model (Duin, Oliveira, and Saffarpour, 2007), which relies on a set of well-defined Key Performance Indicators (KPIs), such as throughput, defect rate, turnover, number of sold products, etc, which are derived from the underlying measures of the simulation model. KPIs describe how the operation process is performing. An initial set of KPIs is given to the user based on the ENAPS Framework (ENAPS Consortium, 2001). The supporting simulation model is extensible, thus it is permissible for an end-user to add more KPIs.

The PRIME Serious Game is based on two game modes:

- **Corporate**. The corporate mode provides an overview of the organization and the necessary information to support high-level strategic decisions. By default, the mode is dominated by the geographical global view. This view only exists in Corporate Mode and the player may extract information on geography, demographics, macro-economics and political systems.
- **Site**. The site mode provides fine granular detail, it is necessary for the player to switch from corporate mode into Site Mode to make localized strategic decisions.

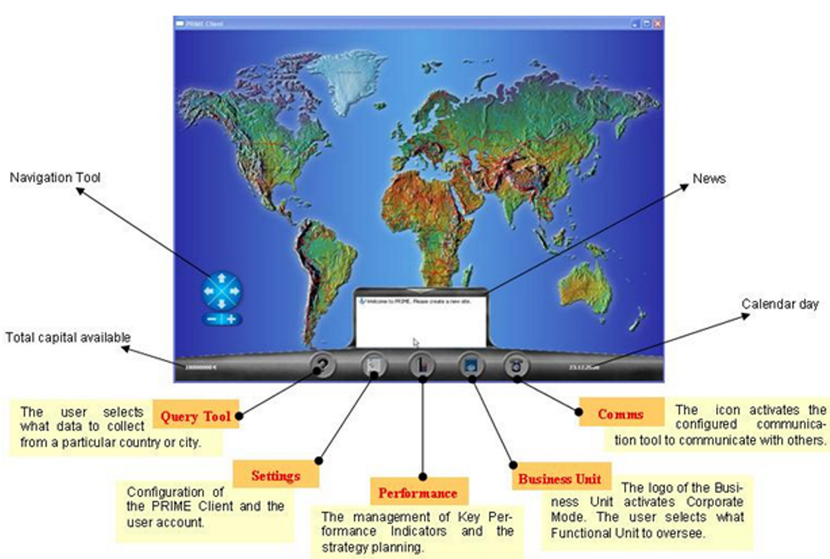
The image of Figure 5 shows an example of PRIME in corporate mode. It is extracted from the prototype corresponding to the final release version of the software, which again is corresponding to a fully working prototype.

Evaluation

The PRIME serious game was evaluated in six end-user companies, from different industrial sectors and with different characteristics:

- Fiat (automotive)
- IAI (aerospace)
- Intracom (telecommunications)
- Kesz (construction)

Figure 5. PRIME in corporate mode



- Lego (toy industry)
- Siemens (software)

The resulting population sample from the 6 end-user companies involved in the evaluation process yielded a total of 50 participants with different gaming experience, which revealed to be detrimental in tackling the learning curve associated to the PRIME serious game. 20% of the participants were players; 50% were seldom players; and 30 were not players. In average they spent 14 hours on PRIME.

The adopted evaluation methodology consisted of three phases:

- **Preparation**: This phase lasted two to three weeks and participants were encouraged to experiment with the PRIME serious game. This was necessary to tackle the learning curve and was done with the support of a moderator located locally within the organization. An underlying aim of PRIME is to facilitate the transfer of experience gained in the serious game to the real world, thus encouraging the building of the learner's repertoire. Consequently, the participant has two forms of preparation: guided-play and free-play. In the former, the participant interacted with the PRIME serious game based on tutorials that are well structured with clear aims for each case. In the case of free-play, the participant has no support or structure to their gaming sessions. During this phase, the participant was requested to log (Training Log) their opinions, difficulties, problems encountered and time spent playing the game.
- **Assessment**. At the end of the preparation phase, the participant was sufficiently familiarized with the PRIME serious game to focus on the learning process rather than any particular issue related to gameplay or game mechanics. The learner was challenged with a crafted "mission" where certain business objectives needed to be achieved within a given business context. During the process, evaluation data was collated in the form of semi-structured interviews, questionnaires, and quantitative data resulting from in-game performance indicators. The collated data forms part of the participant's Mission Log.
- **Analysis**. All the collated data from the two previous phases was analyzed and processed to distil lessons learnt and validate the PRIME serious games.

Only a small minority of participants were mainly players (around 20%), which has an impact on the evaluation data results since there was a correlation between the gaming experience and leveraging the learning curve associated to the PRIME serious game. For the assessment phase, the participants played a crafted mission that included:

- Sales Forecasts
- Project Management
- Product Development
- Production Management
- Logistic Management

Results

Even with a period of one month, the evaluation process could not measure the success of transference of experience into the real world. Consequently, the evaluation focused on usability and user friendliness with some exploration of its potential for achieving learning outcomes.

Since the PRIME serious game was a prototype, it was plagued with some development teething issues that were eventually sorted. The great majority of the learners (66% of the participants) experienced minor problems, but these would not affect their experience. A total of 14% never experienced problems whilst 18% manifested difficulties with the PRIME serious game. However, it should be noted that those experiencing most difficulties were also those individuals who did not have any previous gaming experience. Therefore when confronted with an open system that enabled emergent outcomes, an inexperienced learner would not understand how to proceed. Unlike a behaviorist approach where a learner has a well-defined path, in our case a constructivist approach was taken where there are no clear objectives, but the learner learns by doing. This problem was more self-evident in the Assessment phase where the learner was confronted with a mission with clear objectives, but the non-gamers had serious difficulty in proceeding without the assistance of the moderator. However, after the learner had built their own internal mental model of the process, they would gradually know how to engage and interact with the PRIME serious game.

Another factor to take into account is the fact that PRIME is a serious game on strategic global manufacturing involving management, which is a very complex and interdisciplinary topic. Consequently, PRIME touches most of the aspects of the life of a company, which depending on the size of the enterprise involves the management by a large team of people with different expertise, studies, and competences. Within the population sample, none of the industrial companies were small enough to have a single individual with all the necessary expertise to manage their organization, but neither was this the objective as one design requirement was the alignment of stakeholders by carrying out role reversal.

In addition to the semi-structured interviews and questionnaires, the participants were encouraged to provide comments and feedback in written form, which were important in identifying improvements in the PRIME serious game.

All learners clearly demonstrated, and reported, to enjoy the experience of being engaged with PRIME for learning. However, it was also reported that PRIME was more beneficial when engaged with a group of learners rather than in isolation. For this reason PRIME is located on the borderline between the 3rd and 4th generation of serious games according to Figure 3.

With regards to learning, it was not feasible to have an in-depth assessment that measured the transfer from serious game playing to real-life. However, 76% of the learners reported to acquired new knowledge and gained new understanding that they otherwise didn't have. For this reason, it is not surprising that more than 80% of the learners understood the potential of PRIME as a learning delivery mechanism for global strategic manufacturing.

Recommendations

From the analysis carried out, some guidelines for the improvement of PRIME and serious games in general emerged:

1. As with most sophisticated software, irrespective of how much effort is invested into improving usability, there will always be a learning curve that needs to be tackled by learners. An effective tool to support the leaner is the use of interactive tutorials, in the absence of these; video tutorials are also useful albeit less effective.
2. Learning is more effective in a social and situated context.
3. There is a need for single user experiences where a mission is tailored to very specific tasks. This is the particular case of highly engineered processes, which resemble almost an algorithm that the learner is required to follow. These cases bring the learner a significant return on their time.
4. Learners are partial to simpler missions that require short period of time to be experienced. However, learners recognise that their utility might be limited.
5. A serious game, such as PRIME, would benefit from linking to actual learning objects that provide the theoretical background, thus enabling the learner to have deeper under-

standing of what would be best decisions to make.

6. In the case of PRIME, each session should have a pre-briefing and debriefing to align the learner to what is the purpose of the mission and facilitate the externalization process.

STRATEGIES FOR CORPORATE LEARNING IN SMES

Corporate learning means learning organized by an enterprise – in this case SMEs. The SME sets up the learning objectives and defines the main structure in cooperation with an academic institution. The courses are tailor-made to the enterprise although a part of each course always is generic and therefore applicable independent of a company or even an industrial sector. Examples and cases used are normally based on real data from the enterprises, and assignments address problems the enterprise encounters. Solutions to assignments can be based on company procedures, systems, and governing documentation.

We will describe a case showing an E-Learning approach. Even though the case is from a large company, the approach may be applied also to SMEs. Next some of the challenges experiences in implementing and operating E-Learning will be discussed, and at the end some recommendations are given.

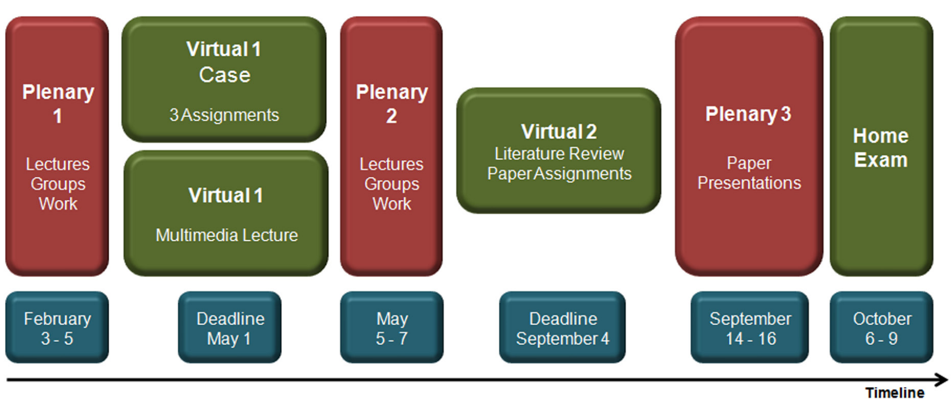
Case Description

The case is a real course developed and delivered in cooperation with a company and a university. The university cooperates with two other universities in different countries bringing a global dimension into the education.

Figure 6 shows an outline of the approach selected in the case. The course was used for training experienced project managers. Participation in each course is limited to 24 persons in order to provide the best possible learning environment. The participants all have more than 10 years of industrial experience from managing projects. Therefore the approach selected is discussion-oriented and highly interactive. The participants learn from each other, and the role of the teacher is more that of a moderator than a lecturer. The teachers are selected both from the company and the cooperating university.

As can be seen the selected approach is based on the hybrid model. There are three plenary sessions, each with three full days of face-to-face learning. The first two plenary sessions contain lectures and substantial group work and discussions. For each day, there are defined learning objectives. An electronic classroom is used for

Figure 6. E-Learning approach used in case



distributing handouts and reading material as well as for discussions amongst the participants and between participants and professors. The third plenary is different, as the participants will present the papers they have written in the virtual 2 assignment. Ample time is allowed to discuss each paper. This stimulates learning from each other and has proved very effective.

The first virtual session is split into two. The first part contains a case description. This is delivered using documents, slide shows, video clips, and interviews with actual stakeholders. From the case, each participant then gets three assignments that they will answer individually. The second virtual session is to write a paper according to normal academic standards. However, there is one important exception. The paper does not need to contain an original research contribution. It will be based on a problem from the company and take the form of a literature survey to find a solution to the problem or a recommendation for the company. This forces the learning to look into literature and discuss findings and alternative solutions. The problem has a "problem owner" in the company. This person serves as a mentor to the learner in addition to the supervision given by the professor. The learners work in groups of about three persons for this assignment. The paper is graded and the participants receive a certificate from the university. As already indicated, the group will have to present the results at the last plenary session.

After the third plenary session, there is a home exam to be answered individually. The exam is published in the virtual classroom and is open for three days. It represents about 6 hours of work, allowing the participant some flexibility in scheduling his or her work. All aids are permitted. The exam has the form of a case description with a number of assignments based on the case. The exam is graded and a certificate is issued by the university.

At the bottom of Figure 6, a typical timeline is indicated. The course is run once every year, and is now into its fourth year.

Challenges

The authors have been engaged in a number of E-Learning courses for different customers and using different approaches. The authors have also studied a number of E-Learning courses available on the web. There are many different solutions, ranging from simple courses presenting a number of slides and running quizzes to advanced multimedia applications.

Some of the most important challenges in developing a successful solution for an SME are discussed below.

The first challenge is to create a stimulating learning environment. This means that the learner is inspired to go on seeking more and more knowledge. A prerequisite for this is that the learning environment is interactive and that the delivery methods applied varies. It is also important that the participant is continuously challenged. The level of education must be adapted to needs of the learner. It should not be too easy or too difficult, and there must be reasonable progress.

Progress of the learner is in itself a challenge. All experience shows that learning is postponed until the very last moment. At the same time research shows that learning is a process maturing over time. It is thus important to break any course into a number of sessions, and to use deadlines and tests to enforce a reasonable progress.

Although progress must be enforced, it should not be so strict that the learner is not allowed any flexibility with respect to time and place for learning. People attending courses delivered through their employer will always have to attend to their normal job in addition. Even though they may use some time for learning during work hours, all experience show that they will have to spend some of their free time for learning as well. For them to succeed and keep up motivation, they should have flexibility to fit learning into their social life in a good way.

For the supplier of the course it is a challenge to respond to the service level demanded from the learner. Using electronic classrooms, one can easily see that questions from participants come in evenings after small children's bedtime and during the weekends. For the teacher it is a challenge to always be available when the participant needs help.

Also technology may represent a challenge. Pure technical questions connected to ICT platforms, security and firewalls may be solved, but requires attention and should not be underestimated. The use of media must also be carefully considered. Video is not always the best solution. The "talking head" may appear boring and reading an article may actually result in better learning. Interactive approaches, such as using serious games, have proved very effective.

Recommendations

For an SME starting to use E-Learning for developing their competence, the authors would like to offer the following recommendations:

1. Develop the courses in cooperation with an academic partner, but make sure the learning objectives are defined by the company.
2. Use hybrid learning based on asynchronous learning for the virtual sessions. Plenary sessions are extremely important for the participants to socialize and to stimulate learning from each other.
3. Base assignments and cases on realistic data from the company.
4. Involve the most qualified people from the company in lecturing practical approaches and company policies. Let the academic partner handle the generic aspects and the theory and try to put company practice into a global generic perspective.
5. Select proven technology for the delivery methods. Such courses are not the battleground for experimenting with cutting edge solutions and for training the employees in learning technology rather than learning content.
6. Provide adequate information on the selected approach to the participants before the first plenary session so that they can enter the learning arena well prepared.

FUTURE RESEARCH DIRECTIONS

In light of the challenges outlined earlier in this chapter, some topics for future research in the area of E-Learning can be identified:

- How to design E-Learning/technology-enhanced learning platforms that are at the same time cost effective in developing and implementing while providing tailored learning for the individual users. This is both a pedagogical and technical problem that is yet unsolved. The pedagogical issues will most likely have to be researched through exploring theories about learning models and concepts that attempt to identify learning needs of individuals (e.g., threshold concept (Meyer and Land, 2003). Technically, tailored learning can probably be realized through the use of mechanisms like serious games with tailored game scenarios, artificial mentors that aid the learners, etc. None of these issues have yet been sufficiently researched to be implemented in mature E-Learning environments.
- Learning progress measurement is another topic that is inherently difficult, but required to develop personalized and tailored learning situations. Determining the slope of the learning curve and presenting the learner with material, situations, and game scenarios that strike the required

balance between a sufficient level of difficulty to both motivate and not scare off and a suitable engagement factor to keep the learner's interest necessitates the ability to measure how the learner performs in various learning situations. Beyond learning quizzes and game scores, very little exists in terms of progress measurement approaches that can be put to use. Future research must look into aspects like cognitive load theory (Sweller, 2005), psycho-physiological measurements, and game scoring principles aimed at uncovering skills and understanding.

- While electronic learning systems have been designed almost exclusively for individual learning processes, i.e., where single learners obtain material, read, solve problems, etc. using their computers, there seems to be a much higher learning potential if at least part of the learning process can be integrated into a larger community of learners. Theories like learning communities (Wenger, 1999) and knowledge ecology (Qvortrup, 2006) have approached this issue from a social sciences perspective, but very little research has been done in terms of how such theories can be exploited to make E-Learning environments more effective. The strength of various on-line forums shows some of the potential, but we have yet to learn how to stimulate such powerful forums in E-Learning applications.

CONCLUSION

Most SMEs are in terms of organizational structure fairly simply designed, with few departments or units. On the other hand, this means that typically, employees are involved in more than one function or discipline required to run the company. In such integrated areas, decisions are taken that have a crucial influence on the competitiveness and profitability of a company. It is therefore mandatory to find adequate learning processes for these organizations to take up new ideas. E-Learning can help in this respect by:

- Reducing the time it takes an employee to acquire the necessary competences
- Changing the learning context rapidly and in response to the real world
- Facilitating knowledge sharing within the organization
- Supporting a soft failure environment where mistakes have no impact on the real world.

One of the most promising, future-oriented concepts of E-Learning is the use of serious games. Studies have shown that gamers have an edge and advantage over those without gaming experience. By introducing games as a tool in the learning process, changing of behaviour and mindset are stimulated and the organization may gradually be transformed into a learning organization.

The PRIME project has experimented with serious games for project business and on-going manufacturing. The project revealed some recommended guidelines for introducing games as learning tools into an organization:

- Irrespective of how much effort is invested into improving usability, there will always be a learning curve that needs to be tackled by learners. An effective tool to support the leaner is the use of interactive tutorials.
- Learning is more effective in a social and situated context.
- There is a need for single user experiences where a mission is tailored to very specific tasks.
- Learners are partial to simpler missions that require short period of time to be experienced.

- A serious game would benefit from linking to actual learning objects that provide the theoretical background.
- Each session should have a pre-briefing and debriefing to align the learner to what is the purpose of the mission and facilitate the externalization process.

For the delivery of E-Learning content, the concept of hybrid learning is recommended. A successful implementation should follow the following guidelines:

- Develop the courses in cooperation with an academic partner.
- Use hybrid learning based on asynchronous learning for the virtual sessions.
- Base assignments and cases on realistic data from the company.
- Involve the most qualified people from the company in lecturing practical issues.
- Select proven technology for the delivery methods.
- Provide adequate information on the selected approach to the participants before the first plenary session.

There is a need for further research in three directions:

- Design of cost effective technology-enhanced learning platforms that provides tailored learning for the individual users.
- Learning progress measurement.
- Integration of the learner process into larger communities of learners.

REFERENCES

Argyris, C., & Schön, D. (1978). *Organizational Learning: A theory of action perspective*. Reading, MA: Addison-Wesley.

Aspin, D., & Chapman, J. (2000). Lifelong learning: concepts and conceptions. *International Journal of Lifelong Education*, *19*(1), 2–19. doi:10.1080/026013700293421

Berge, Z.L. (1998). Barriers to online teaching in post-secondary institutions. *Online Journal of Distance Education Administration, 1(2)*.

Carliner, C. (2002). Tips for Conducting Level 1 Evaluations of Asynchronous e-Courses. *American Society for Training and Development publication What Works, March*.

Dale, E. (1969). *Audiovisual Methods in Teaching*. New York: The Dryden Press.

Duin, H., Oliveira, M., & Saffarpour, A. (2007). A Simulation Model for Virtual Manufacturing Environments for Serious Games. In Proceedings from *13th International Conference on Concurrent Enterprising: Concurrent Innovation: An Emerging Paradigm for Collaboration and Competitiveness in the Extended Enterprise.*

ENAPS Consortium. (2001). *ENAPS Indicator Model, Version 1.0*. Bremen, Germany: PRODUTEC Ingenieurgesellschaft mbH.

Flanagan, J. (1954). The Critical Incident Technique. *Psychological Bulletin*, *51*(4). doi:10.1037/h0061470

Hall, B., & Snider, A. (2000). Glossary: The hottest buzz words in the industry. In Schreiber, D. A., & Berge, Z. L. (Eds.), *Distance training: How innovative organizations are using technology to maximize learning and business objectives*. San Francisco: Jossey-Bass.

Herring, T., & Job, R. (2008). The Next Generation of eLearning: Instructional Alternate Reality Games. *Online Forums*. Retrieved October 15, 2009 from http://www.elearningguild.com/olf/olfarchives/index.cfm?id=387&action=viewonly.

Hughes, J., & Atwell, G. (2003). A Framework for the Evaluation of E-Learning. In *Proceedings of European Seminars – Exploring Models and Partnerships for E-Learning in SMEs*. Retrieved February 4, 2004 from http://www.theknownet.com/ict_smes_seminars/papers/Hughes_Attwell.html.

Kok, W. (2004). *Facing the Challenge: The Lisbon strategy for growth and employment, report from the High Level Group*, available at http://europa.eu.int/growthandjobs/pdf/kok_report_en.pdf.

Kolb, D. A. (1984). *Experimental Learning – Experience as The Source of Learning and Development*. Upper Saddle River, NJ: Prentice Hall.

Kommers, P. A. M., Grabinger, S., & Dunlap, J. C. (1996). *Hypermedia Learning Environment: Introduction Design and Integration*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lefrere, P. (2007). Technology to support distance education. In proceedings from *Strategies for Global Manufacturing: a European view of IMS*. Available from ftp://ftp.cordis.europa.eu/pub/ims/docs/4-6-lefrere.pdf.

Leonard-Barton, D. (1995). *Wellsprings of Knowledge – Building and sustaining the sources of innovation*. Cambridge, MA: Harvard Business School Press.

March, J. (2006). Ideas as Art. *Harvard Business Review*, (October): 83–89.

Meyer, A., & Rose, D. H. (1998). *Learning to Read in the Computer Age*. Cambridge, MA: Brookline Books.

Meyer, J. H. F., & Land, R. (2003). Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines. In Rust, C. (Ed.), *Improving Student Learning: Improving Student Learning Theory and Practice – Ten Years On*. Oxford: Oxford Centre for Staff and Learning Development.

Oliveira, O., Andersen, B., Oliveira, A., & Rolstadås, A. (2006). The Impact of Serious Games in the European Industry. In proceedings from the *10th Workshop on Experimental Interactive Learning in Industrial Management - IFIP WG5.7 SIG*, IFIP.

Polanyi, M. (1967). *The Tacit Dimension*. Garden City, NY: Doubleday.

Prensky, M. (2001). *Digital Game-Based Learning*. New York: McGraw Hill.

Qvortrup, L. (2006). *Knowledge, Education and Learning: E-Learning in the knowledge society*. Copenhagen: Forlaget Samfundslitteratur.

Rolstadås, A. (2002). An Example of an E-Learning Solution for an International Curriculum in Manufacturing Strategy. In Proceedings from *e-2002 Conference*.

Rolstadås, A., & Hussein, B. (2002). Hybrid Learning in Project Management – Potentials and Challenges. In Proceedings from *PMI Research Conference*, PMI.

Romiszowski, A. (2004). How's the E-Learning Baby? Factors Leading to Success or Failure of an Education Technology Innovation. *Educational Technology*, *44*(1).

Rose, C., & Nicholl, M. (1997). *Accelerated Learning for the Twenty-First Century*. New York: Delacote Press.

Sage Learning System. (2002). E-Learning Evaluation. Retrieved July 5, 2003 from http://www.sagelearning.com/papers/Evaluation%20Forms.pdf.

Sawyer, S., & Tapia, A. (2005). The sociotechnical nature of mobile computing work: Evidence from a study of policing in the United States. *International Journal of Technology and Human Interaction*, *1*(3), 1–14.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.

Schrage, M. (2000). *Serious Play – How the Worlds Best Companies Simulate to Innovate*. Cambridge, MA: Harvard Business School Press.

Schwesig, M., Rolstadås, A., & Thoben, K.-D. (2005). An E-Learning experiment in manufacturing strategy. In proceedings from *8th IFIP World Conference on Computers in Education*, IFIP.

Snowden, D. J., & Boone, M. (2007). A Leader's Framework for Decision Making. *Harvard Business Review*, (November): 69–76.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In Mayer, R. E. (Ed.), *Cambridge handbook of multimedia learning*. Cambridge, MA: Cambridge University Press.

Urdan, T. A., & Weggen, C. C. (2000). *Corporate E-Learning: Exploring a new frontier*. San Francisco, CA: WR Hambrecht & Co.

U.S.I.G. (2003). Usability Special Interest Group: Usability checklist for E-Learning. Retrieved December 1, 2003 from http://www.stcsig.org/usability/resources/toolkit/E-Learning-checklist.doc.

VandenBos, G., Knapp, S., & Doe, J. (2001). Role of reference elements in the selection of resources by psychology undergraduates. [from http://jbr.org/articles.html]. *Journal of Bibliographic Research*, *5*, 117–123. Retrieved October 13, 2001.

Wenger, E. (1999). *Communities of Practice: Learning, Meaning and Identity*. Cambridge, UK: Cambridge University Press.

This work was previously published in E-Business Issues, Challenges and Opportunities for SMEs: Driving Competitiveness, edited by Maria Manuela Cruz-Cunha and João Varajão, pp. 46-64, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 7
Categorization of Losses across Supply Chains:
Cases of Manufacturing Firms

Priyanka Singh
Jet Airways Limited, India

Faraz Syed
Shri Shankaracharya Group of Institutions, India

Geetika Sinha
ICICI Lombard, India

ABSTRACT

Supply chain loss can occur during transit and storage, leading to unnecessary inefficiencies. The literature details much of the traditional losses, albeit descriptively and for developed economies. Through several case studies conducted on the Indian manufacturers and retailers, this case study discusses the losses specific to supply chains operating in developing economies that are difficult to control and prevent even with contemporary enabling technologies such as RFID. This chapter also suggests some possible measures to counter such losses, so as to increase the efficiency and enhance the resilience of the supply chain. An understanding of these losses and their possible mitigation through improved flows, reduced inventory, and reduced manpower, can equip firms for better supply chain risk and productivity management.

INTRODUCTION

Losses are quite common during transit and can occur due to various reasons including theft, tampering and spillage. The type of loss also depends upon the nature of product. Bulk products are more prone to losses than packaged products.

DOI: 10.4018/978-1-4666-1945-6.ch007

Similarly, the type of transport used also determines the losses that would occur in the supply chain. Traditionally, firms allow some tolerance limit for such losses. However, any loss during the supply chain is still a loss. This is truer with increased competition and emphasis on all around cost-cutting. Lets us consider the example of coal transport. If there are 100 trucks supplying coal to a steel plant everyday and each truck can deliver 10 tonnes of coal. A tolerance limit of 1% means there would be a loss of 10 ton of coal every day. Considering the prevailing price of $50 per ton of coal, the total loss would be around $500 everyday and $0.15 million in terms of annual cost. This is a crude estimate.

Different supply chains are prone to different types of losses. For example, supply chains of perishable goods are prone to losses due to perishability of food products. Bulk products are prone to adulteration with lower quality bulk material as well as theft and spillage. Packaged goods are prone to damage during transit and sometimes intentional tampering is done, such as in crockery so that it will be sold at substantially lower prices.

Here, two case studies are reported which were conducted to identify these supply chain losses.

CASE 1: LOSSES IN SUPPLY CHAIN OF BULK PRODUCTS: COAL

The Indian energy sector is largely dependent on coal as the prime source of energy. After the Indian independence, a great need of coal production was felt in first five year plan. In 1951 a working party for the coal industry was setup, which suggested the amalgamation of small and fragmented producing units. This led to the idea of a unified coal sector [1]. Coal is an essential ingredient in various industries such as steel industry, thermal power plants, hydro-electric power plants, manufacturing industry, and cement industry.

Coal Mining in India

In the pre-nationalized era coal mining was controlled by private owners, and suffered from their lack of interest in scientific methods, unhealthy mining practices and sole motive of profiting. The miner lived in sub standard conditions as well. In 1956, the National Coal Development Corporation was formed with 11 collieries with the task of exploring new coalfields and expediting development of new coal mines[1]. The objective of nationalization was the conservation of the scarce coal resources, particularly coking coal. Later the name of NCDC was changed to Coal India Limited (CIL). Coal India Limited is a Schedule ‘A’ ‘Navratna’ Public Sector Undertaking under Ministry of Coal, Government of India, with Headquarters in Kolkata, West Bengal1. CIL is the single largest coal producing company in the world and the largest corporate employer in the country with manpower of 409,332 (as on 1 July 2009). With proven coal reserves of 105.82 Billion Tonnes out of total reserves of 267 Billion Tonnes (as on 1 April 2009) Coal India plays a pivotal role in Indian energy scenario. The Mission of Coal India Limited is to produce the planned quantity of coal, efficiently and economically with due regard to safety, conservation and quality[2]. Coal India is a holding company with seven wholly owned coal producing subsidiary companies and one mine planning & Consultancy Company. It encompasses the whole gamut of identification of coal reserves, detailed exploration followed by design and implementation and optimizing operations for coal extraction in its mines[2]. The producing companies are Eastern Coalfields Limited (ECL), Sanctoria, West Bengal; Bharat Coking Coal Limited (BCCL), Dhanbad, Jharkhand; Central Coalfields Limited (CCL), Ranchi, Jharkhand; South Eastern Coalfields Limited (SECL), Chattisgarh; Western Coalfields Limited (WCL), Nagpur, Maharashtra; Northern Coalfields Limited (NCL), Singrauli, Madhya Pradesh; and Mahanadi Coalfields Limtied (MCL), Sambalpur, Orissa; The consultancy company is Central Mine Planning and Design Institute Limited (CMPDIL), Ranchi, Jharkhand. North Eastern Coalfields (NEC) a small coal producing unit operating in Margherita, Assam is under direct operational control of CIL.

Types of Coal

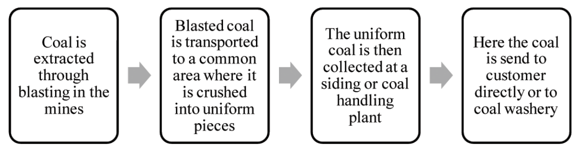
There are various types of coal and coking coal is considered the best variety of coal in India. These coals, when heated in the absence of air, form coherent beads, free from volatiles, with strong and porous mass, called coke. There is also semi-coking coal which when heated in the absence of air form coherent beads not strong enough to be directly fed into the blast furnace. Such coals are blended with coking coal in adequate proportion to make coke. The third variety NLW coking coal is not used in metallurgical industries and is also not acceptable for washing in coal washeries because of its high ash content. This coal is used for power utilities and non-core sector consumers. The coal with high ash content is washed in coal washeries. The washing process reduces the ash content of coal the coal becomes suitable for being used in steel and other industries. The coal without coking properties is known as non-coking coal. The byproducts of coal washing and beneficiation process are called middling and can still be used for various purposes. After removing clean coal and middling from the coal beneficiation processes the leftover is thrown away as rejects. CIL coke/ LTC coke is a smokeless, environment friendly product of coke oven plants.

The gradation of non-coking coal is based on Useful Heat Value (UHV), the gradation of coking coal is based on ash content and for semi coking / weakly coking coal it is based on ash plus moisture content, as in vogue as per notification3. The various grades of coal are listed in Table 1.

The suitability of coal for any particular use depends upon its grade. Table 2 shows the suitability of coal for any particular use.

Mining and Loading of Coal

The process of mining is simple but in practice it is very-very complex. The coal is extracted through blasting and the blasted coal is transported to a crusher, where it is crushed into uniform pieces. The amount of coal transported is crudely recorded in terms of the no. of trips made to the crusher by the truck / dumper and the capacity of the dumper. The coal is then collected at a siding or a coal handling plant. Here the coal is weighted at the weigh bridge and its weight is recorded. The amount of coal received from the mine (no. of trips x capacity of truck / dumper) is then matched with the quantity recorded at weighment bridge.



If impurity content in coal is more than the desired amount than this coal is sent to Coal Washing Plant. The main purpose of this plant is to remove impurity. Once this impurity is removed they are sent to siding. Siding is a place usually near railway stations, where coal from the mine is collected. To avoid any kind of dust disposition, water is spread over the coal. Here also different types of losses occur. If coal is transported to government sector or public sectors, sometimes inferior quality coal is added to require quality at the agreed rate.

From siding coal is transported to different customers and industries through train or truck. When coal is being dispatched, representatives of buyer and seller are present at the delivery point to make sure that the agreed quantity is being dispatched. From here the coal is send to the customer. The major customers of coal include Cement industry, Fertilizer and chemical industries, Steel industry, and Thermal power stations.

Several losses take place during this process. Before the weight is recorded, the losses are basically due to spillage during transportation and minor theft by the miners themselves. Major losses occur due to theft conducted by authorities themselves. Suppose the coal is transported by a truck / dumper from the blasting site to the hopper. If a truck / dumper of 35 tonnes capacity make 10

Table 1. Coal gradation and types

GRADES OF COKING COAL	
Grade	**Ash Content**
Steel Grade –I	Not exceeding 15%
Steel Grade -II	Exceeding 15% but not exceeding 18%
Washery Grade -I	Exceeding 18% but not exceeding 21%
Washery Grade -II	Exceeding 21% but not exceeding 24%
Washery Grade -III	Exceeding 24% but not exceeding 28%
Washery Grade -IV	Exceeding 28% but not exceeding 35%

GRADES OF NON-COKING COAL			
Grade	**Useful Heat Value (UHV) (Kcal/Kg) UHV= 8900-138(A+M)**	**Corresponding Ash% + Moisture % at (60% RH & 40^{O} C)**	**Gross Calorific Value GCV (Kcal/ Kg) (at 5% moisture level)**
A	Exceeding 6200	Not exceeding 19.5	6454
B	5600 to 6200	19.6 to 23.8	6049 to 6454
C	4940 to 5600	23.9 to 28.6	5597 to 6049
D	4200 to 4940	28.7 to 34.0	5089 to 5597
E	3360 to 4200	34.1 to 40.0	4324 to 5089
F	2400 to 3360	40.1 to 47.0	3865 to 4324
G	1300 to 2400	47.1 to 55.0	3113 to 3865

GRADES OF SEMI-COKING AND WEAKLY COKING COAL	
Grade	**Ash + Moisture Content**
Semi coking grade –I	Not exceeding 19%
Semi coking grade –II	Exceeding 19% but not exceeding 24%

GRADES OF NEC COAL		
Grades	**UHV (Kcal/Kg)**	**Corresponding Ash% + Moisture %age**
A	6200-6299	18.85 – 19.57
B	5600 – 6199	19.58 – 23.91

trips then the total quantity transported should be 350 tonnes. If these trucks are overloaded beyond their maximum capacity then the quantity transported would be more than 350 tonnes although the recorded total will still be 350 tonnes. This extra tonnage transported can now be stolen or sold to third party without one being charged for theft.

Theft can also be conducted by the dumper operators who may dump the coal in a different area rather than the stipulated place. Truck drivers either dump some portion of coal (which they carry for transporting of coal) in their area or they use it as mode of making payment to those where they take their meal.

Even at the dispatch locations, the nexus with the security guards and other related people increases the possibility of theft. A truck of coal would cost around Rs. 75000 to the company and if a company loses 10 trucks every month, then such loss would account to Rs. 10 million in annual costs. The total cost due to pilferage is very high to a firm thereby warranting adoption of a

Table 2. Types of coal and their industry use

Industry	Types of coal required
Cokeries/ coke oven plants	Coking and semi-coking coal
Steel making	Coking and semi-coking coal, direct feed and washed; blendable coal; low ash% ASSAM and Ranigunj coal
Sponge iron industry	Non-coking coal of high Initial Deformation Temperature
Domestic fuel making	Semi-coking and non-coking coal; Middlings & rejects of Washeries
Special Smokeless fuel	Semi-coking coal of index 8-10
Power sector	Non coking-coal; middlings of coking coal washeries; washed coal of non-coking coal washeries
Cement sector	Non-coking coal; middlings of coking coal washeries
Steel casting	Non-coking coal
Domestic use, Hotel and Restaurants	Non-coking coal; CIL coke/LTC coke

suitable technology that can improve supply chain visibility. RFID can help reduce theft at the point of dispatch by automating the security operations. Only those trucks that have RFID readers would be allowed to go inside the mines and come out of the mine thus preventing unauthorized entry and exit of trucks.

Transportation of Coal

Coal industries mainly use railways and roadways for transporting coal within the county. India also exports coal in negligible amount through waterways. India's transport sector is large and diverse; it caters to the needs of 1.1 billion people. In 2007, the sector contributed about 5.5 percent to the nation's GDP, with road transportation contributing the lion's share. Good physical connectivity in the urban and rural areas is essential for economic growth. Since the early 1990s, India's growing economy has witnessed a rise in demand for transport infrastructure and services. However, the sector has not been able to keep pace with rising demand and is proving to be a drag on the economy. Major improvements in the sector are required to support the country's continued economic growth and to reduce poverty[4].

Generally bulk products within the country are transported using railways or roadways. Indian Railways is one of the largest railways under single management. It carries some 17 million passengers and 2 million tonnes of freight a day in year 2007 and is one of the world's largest employers[4]. The railways play a leading role in carrying passengers and cargo across India's vast territory. However, most of its major corridors have capacity constraint requiring capacity enhancement plans. Railways provide special types of Bogies for transporting specific material. Steel rods, bars and blooms are generally transported on the chassis of a bogie. Coal is transported in open bogies with hopper. Grains etc. are transported in closed bogies. For transporting petroleum products, the bogies are designed in cylindrical shape. Railways also conduct container transport across the country. Railways are generally preferred for long distance transportation. Indian Railways also provide facility for captive siding where companies can load / unload products transported to their dispatch / unload section.

For short distance, roadways are dominant mode of transportation. They carry almost 90 percent of the country's passenger traffic and 65 percent of its freight. The density of India's highway network -- at 0.66 km of highway per square kilometer of land – is similar to that of the United States (0.65) and much greater than China's (0.16) or Brazil's (0.20)[4]. However, most highways in India are narrow and congested with poor surface quality, and 40 percent of India's villages do not have access to all-weather roads.

Transportation provides mobility to heavy and bulky raw material to basic industries like iron and steel. It leads to economic prosperity and infrastructural development of any economy. Core industries like iron and steel, coal industry, cement industry etc depends heavily on various means of transportation for acquiring its raw material and for making its finished goods available in the market.

Losses during Transportation of Coal from Coal Handling Plant and Coal Washery

The coal being transported to the customer is prone to several types of losses such as theft, spillage and adulteration. These losses are mentioned below:

Spillage

Spillage is one of the major causes of loss during transportation although the loss through spillage is negligible. Because of poor road conditions, potholed roads, speed breakers, narrow pathways etc. the losses take place. However, many-a-times the drivers are quite cautious to not allow such spillage to take place. This allows them to steal the coal and mark it as a loss due to spillage.

Pilferage of Goods / Materials during Transit

Most manufacturing firms in India receive goods in bulk form their raw material suppliers. These materials may include materials such as coal, ore, sand, bricks and petrol / diesel. Firms account for 2-3% of received material as loss during transit and consider this loss as normal. However, the transporters and transport operators take advantage of this allowance to conduct pilferage activities along the route from supplier to customer. Even this 2-3% loss could cost a firm millions of Rupees lost. For example, for a truck of coal transporting 30 tons of coal, pilferage of 1 ton of coal would amount to around 3% of loss. One ton of coal costs around Rs. 1500 to the firm. A firm receiving 30 trucks per day of coal will have to bear around Rs. 45000 everyday and around 1.5 million rupees annually. RFID technology coupled with GPRS technology thus holds a great potential for these firms, but even with RFID it is not easy to contain pilferage of goods during transit. This is because transport operators are known to cause damage to such tracking equipment and show erratic behavior when they are tracked along the supply chain. Moreover, even if their location is tracked, their nexus with naxalites, police and local junta make the task of tracking more difficult for the firm.

Theft and Nexus

This is perhaps the biggest cause of loss during transportation. Truck drivers dump some portion of coal (usually this is equal to the 3-4% tolerance limit of loss allowed during transportation) in their area and use it as a mode of making payment to the highway restaurants where they take meal. Sometimes truck drivers form nexus with people who help them to steal coal during transportation.

Sarda Energy and Minerals Limited, Siltara, Raipur (India) took steps to reduce such coal loss by covering the truck with tarpaulin and sealing it with 50-70 lead seals. A broken lead seals is an indication of tampering done with the sealing and a possible adulteration or theft, although it is likely that the seal may broke during transportation itself. Each seal costs around Rs. 10 thus costing the company Rs. 700 per truck per trip as well as 0.5-1.0 hr taken to seal each truck. The seals have reportedly reduced theft to 0.1%.

Another way to know whether theft has taken place is by recording the total transportation time and comparing it with average transportation time. If the total transportation time is exceedingly large as compared to average transportation time then it is a likely indication of theft.

Adulteration

Adulteration refers to mixing or adding of inferior quality or sub-standard material with the actual product. This happens in the case of coal also. Sometimes stone (usually bigger in size and black in colour) are added to coal to increase its weight. This generally happens when coal is supplied to government sectors. Sometimes, truck drivers replace coal with equivalent quantity of water so as to take care of discrepancy during weighment. These can however be easily revealed through chemical analysis of coal sample.

CASE 2: LOSSES IN SUPPLY CHAIN FOR MANUFACTURED GOODS: STEEL

Simplex Group of Industries is one of the leading Indian manufacturers of varied and diversified Engineering products, Castings and Equipments for all core industrial segments for over 40 years. With such a diversified Engineering experience and technical know-how – coupled with state of the art Design, Manufacturing and NDT facilities, Simplex is now synonyms with quality products worldwide. Simplex Group of Industries has six plants situated in and around Bhilai, Chhattisgarh and Madhya Pradesh. Simplex, as a multifaceted organization, has also acquired a prestigious branding in the heavy industry segment as an Equipment builder, Technology supplier and a Turnkey contractor. With the backing of more than 3000 strong and experienced professional workforce of highly skilled workmen and Engineers including 350 no. of advanced and sophisticated machine tools, meeting customer demands with quality and timely deliveries has become a culture in Simplex. With a well established network of offices in all metropolitan cities and resident representatives in all industrial segments, we are indeed growing by the day[5].

Simplex Castings engaged in manufacture of Heavy Engineering Castings in various grades (as cast supply, machined & assembled condition), for core industrial sector (both Domestic & Global level) like Steel, Power (Thermal, Hydel, Wind), Rail, Road, Mines, Cement, Oil, Defence, Sugar, Construction Earth Moving and other specialized areas. Simplex also contributes to industrial development with its highly experienced and professional staff. Given below is their manufacturing capacity[5].

Manufacturing Facilities at Simplex Castings

Grey Cast Iron Castings: Single piece weight from 05 Kgs. To 32000 Kgs.
Ductile / S.G Iron Castings: Single piece weight from 05 Kgs. To 15000 Kgs.
Cast Steel Castings: Single piece weight from 10 Kgs. To 27000 Kgs.
Stainless Steel Castings: Single piece weight from 350 Kgs. To 12000 Kgs

Other facilities include:

Pattern making: Fully equipped with wood working machines pattern shop with experienced pattern makers.
Heat Treatment: In-house facilities for stress relieving, normalizing, quenching etc. to achieve the properties of castings as per technical delivery conditions.
Machining: Equipped with 300 sophisticated machine tools comprising of Horizontal and Vertical Boring Machines, Milling Machines, CNC Lathe, Radial Drilling Machines, Plano Millers and Tool Planners.
Testing Facilities: At Simplex Castings foundries are equipped with versatile chemical and

physical Dye Penetrating Ultrasonic testing and all destructive and non-destructive facilities including spectrometers, microscope with cameras, cobalt 60 source for radiography, hydrogen analyzer for steel etc.

Basic Design and Detailing Facilities: For carrying out detailed design work, Simplex comprises full-fledged Design Department with CAD facilities, scanners and plotters where we design equipment with minimum possible weight.

Quality Standards: The products of Simplex Castings are accepted by inspection agencies like AAR, LLOYDS, RITES, RDSO, SGS, BUREAU VERITAS and other such international inspection agencies.

The product range at Simplex Castings include unalloyed plain carbon steel casting, alloy steel casting, 1.5% mn steel casting, hi-tensile steel casting, corrosion steel casting, ductile steel casting in d-2 grade, ductile iron, stainless steel casting, cast iron, heat-resistant cast iron, Ni-resistant cast iron.

Supply Chain at Simplex Castings

Raw Materials

In steel casting the two major raw materials are Sponge Iron and Steel Scrap. The sponge iron costs around Rs. 14 per Kg (App. $0.30 per Kg) and the major suppliers are Raipur Alloys, Jindal steel, Monnet Ispat and Topworth Steel. The sponge iron is transported through trucks capable of 24 tons capacity. Steel scrap costs around Rs. 18 per Kg (App. $0.40 per Kg). The major suppliers of steel scrap are Bhilai Steel Plant and Rourkela Steel Plant which transport the scrap via truck capable of carrying 16-18 tons. Some steel scrap is also imported from South Africa and Middle East using waterways to Vishakapatnam Port. From Visakhapatnam the steel scrap is transport by railways and roadways. Around 50-60 Kgs of steel scrap is stolen every week or two.

Steel Casting Process

Figure 2 shows the steel casting process at Simplex Casting.

Marketing and Supply Network of Simplex Castings

Besides its full fledged Marketing Office at all the four metropolitan cities of India, Simplex Castings have representation in all major Industrial Sectors (both National & International level) too. Today Simplex is a rising International Star with a rapidly growing Export Market and a commitment that places customer above all else. Simplex exports castings in as *cast & finish machined condition* to countries like U.S.A., Japan, France, Italy, Australia, and Egypt. Presently their export constitutes 30% of their annual turnover.

Losses during Transportation

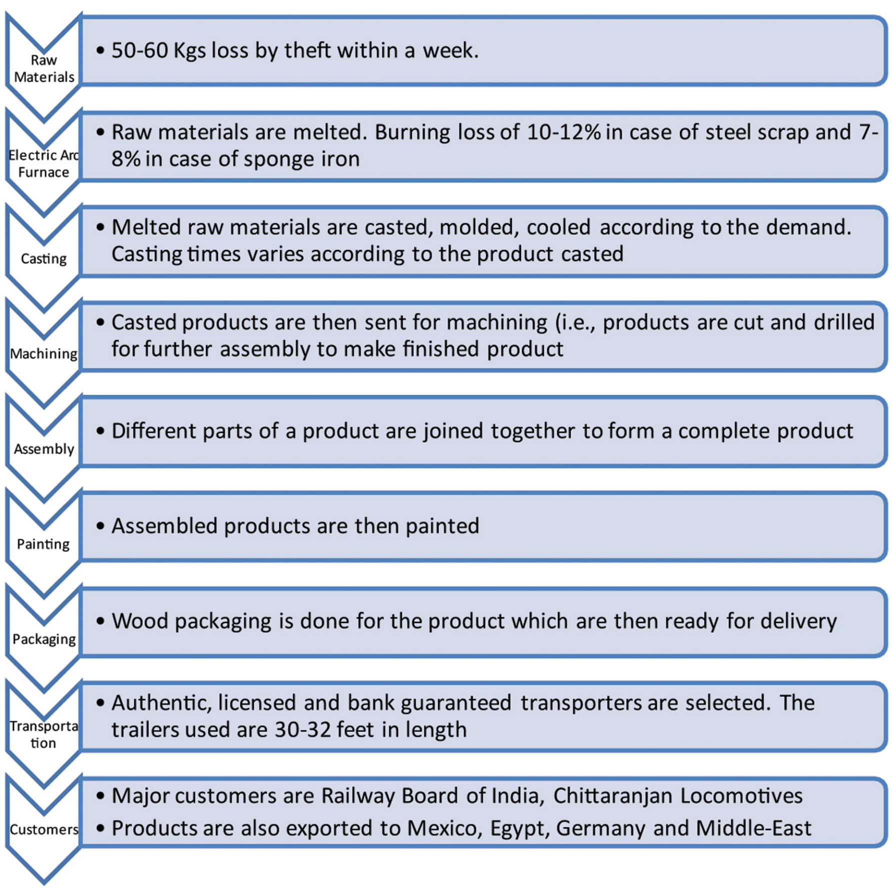
Both direct and indirect losses occur due to supply chain. If there is a delay in supply of raw materials, the production has an adverse effect which in turn affects dispatch of finished products. Delay in transportation also occurs due to strikes and accidents.

Theft usually occurs for small accessories while transportation. It is difficult to steal products that are bigger in size. The accessories that are included with these bigger products are however prone to theft. Moreover, it is difficult to trace the theft of these accessories.

If there is a delay in the time given by customers for delivery of goods LD (Late Delivery) has to be paid by the company which is 0.5% of the total value of the product which is calculated on per month basis. Company has to bear any kind of loss due to accidents while transportation, though the goods are insured but are not 100%, so again there is loss of manpower, energy, raw materials. Due to delay in delivery of products, there is also delay in payment.

Figure 1. Processes at Simplex Castings

Figure 2. Supply chain and steel casting process at Simplex Castings



Another difficulty lies is difficulty in tracking the movement of goods as there is no proper tracking method. Companies are now introducing Global Positioning System on transportation vehicles for tracking them during transportation.

Interference Problem with Steel

RFID tags are vulnerable to interference with steel. Some companies have suggested placing tags by sticking it to the end point of the steel. However, the rugged conditions of manufacturing plant render this possibility useless. Tags must be suitable for bearing wear and tear and hence must be placed snuggly on the bloom. Many RFID vendors have designed tags for steel equipment and these can be used on a trial basis for tagging blooms. These tags can be fitted to the blooms using a screw. However, blooms required for rail mill and plate mill cannot be tagged as they are re-heated / re-melted for further processing.

Reading Distance

In steel plants, at least 6-8 meter distance is required for placing RFID readers. RFID tags can comply with this distance, however, they are costly as compared to ordinary tags and a single tag can cost up to around Rs. 60-100.

CONCLUSION

The current cases discusses about the losses during transportation of bulk products (Coal) and bulky products, namely, steel casts. In case of bulky products, there are only a few reports of theft during the transportation of these products because they are very bulky in nature. The losses commonly occur in such supply chains due to delay in transportation, accidents and strikes. In case of bulk products, however, theft is quite common. The unholy nexus among local people and truck operators is a major cause of losses across supply chains. The remedial measures involve acceptance of Global Positioning System along with RFID technology. Through this technology people can be tracked and if there is unnecessary delay in the Supply Chain they can be questioned. But still there is a long way to go in preventing theft across the supply chain.

ENDNOTES

[1] http://www.coalindia.in/Company.aspx?tab=3
[2] http://www.coalindia.in/Company.aspx?tab=0
[3] http://coal.nic.in/point4.html
[4] http://web.worldbank.org
[5] www.simplexcastings.org

This work was previously published in Cases on Supply Chain and Distribution Management: Issues and Principles, edited by Miti Garg and Sumeet Gupta, pp. 373-386, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 8
Collaborative Demand and Supply Planning Networks

Hans-Henrik Hvolby
Aalborg University, Denmark

Kenn Steger-Jensen
Aalborg University, Denmark

Erlend Alfnes
Norwegian University of Science and Technology, Norway

Heidi C. Dreyer
Norwegian University of Science and Technology, Norway

ABSTRACT

The focus of manufacturing planning and control has gradually expanded from (in-house) production activities towards all manufacturing and logistic activities in the supply chain. Planning of in-house operations is still very important, but the trends towards increased use of outsourcing and mass customisation require that customers and suppliers are able to exchange information frequently to cut down costs and lead time while quickly adapting their manufacturing and logistics operations to market/customer requirements. Many vendors offer systems to plan and control in-house operations, whereas only a few large vendors (such as Oracle, SAP and I2) offer supply chain planning systems. This limits the ability for SMEs to exploit the supply chain planning options. This chapter discuss current supply chain planning solutions and presents a more simple and adaptive concept to be used in both SMEs and larger enterprises. The research presented in this chapter is funded by the EU Union via the EmpoSME, ValuePole projects, and by the Research Council of Norway via the SFI Norman project.

INTRODUCTION

Most manufacturing companies face strong competition and continuous changes in market and customer requirements. Planning are frequently affected by the actions of suppliers and customers in their supply networks such as changes to orders that are already in production or re-planning caused by lack of materials or resources. Often, decisions have to be made without having a complete, real time overview of possible options and associated consequences. The result is typically excess inventories, too long lead times, too low customer satisfaction and poor resource utilisation.

DOI: 10.4018/978-1-4666-1945-6.ch008

Supply networks are dynamic and hard to define, and a single company is often part of several different supply networks simultaneously. An increasing level of customisation combined with demands for high quality, low costs, short and precise delivery times, and high flexibility represents a significant challenge to managing operations in networks.

Although these challenges will apply to companies of all sizes, they are particularly true for SMEs due to the following characteristics:

- The network are typically non-hierarchical
- They have limited staff available for specialist roles in planning and decision making
- They have limited resources available to invest in Advanced Planning Systems
- They require flexible tools which let them exploit the advantage of typically being flexible and easier to manage than larger enterprises

Observation shows that information exchange is often limited to order placements without any kind of information visibility or other communication between network partners. Thus network effects of individual decisions are often not possible to neither intercept nor predict. This implies that planning and control at any network partner is currently executed with incomplete information about status among the other network partners and without the possibility to see the full consequences of decisions being made. Additionally, the planning and control task will vary with regards to scope and complexity challenging the traditional planning and control approaches, and existing methodologies, tools and knowledge.

The paper briefly discusses three major areas for manufacturing in non-hierarchical networks. After this a more simple and adaptive alternative approach to existing APS solutions is presented, and finally the approach is positioned in a broader perspective titled "Work Bench Concept" to be used in both SMEs and larger enterprises.

COLLABORATION IN NON-HIERARCHICAL NETWORKS

Typical for SMEs is that they operate in non-hierarchical networks, characterised by power being distributed among members, and the absence of one or several dominant actors that dictate plans or impose a centralised planning perspective (Harland et al., 2001). In such networks each member participates in multiple supply chains and all members are more or less equal in status and therefore no member has the power to dictate the others (Jagdev and Thoben, 2001). The core part of each network might form what could be titled as a Virtual Enterprise to share skills or core competencies and resources in order to better respond to business opportunities.

Coordination and collaboration between the companies is vital in such networks. Collaboration refers to the activities and environment related to the "joint planning and execution of supply chain activities" (Ayers, 2006), and is therefore an essential element in planning activities in a network perspective. Collaboration is using cooperative efforts in order to meet mutual goals, exchanging information, developing improvement in partnership (Ayers, 2006). A manufacturing network is fully coordinated when all decisions are aligned to accomplish global system objectives (Sahin and Robinson 2002); of course when this occurs decisions have already crossed the company's boundaries, meeting articulated and complex contexts (Danese et al., 2004). Several collaborative models for coordination networks activities have been developed. The aim of models such as collaborative planning, forecasting and replenishment (CPFR), vendor managed inventory (VMI) and automated replenishment programs (ARP) is to achieve seamless inter-organisational interfaces by specifying control principles and operations models for the flow of materials and information (Holweg et al., 2005).

Collaboration in the network is decentralised and each actor performs individual planning, with the aim of optimising operations in a local perspective. Companies in non-hierarchical networks face challenges that make collaboration difficult. Firstly, the networks themselves are dynamic and hard to define, and a single company is often part of several different manufacturing networks simultaneously. The focal company and other network members frequently face situations with conflicting interest and trade-off situations. A typical scenario is when orders from different networks compete for scarce resources and deadlines and delivery times are short. Frequently, these requirements cannot be fulfilled because there is a lack of information about the current state of the production and network processes. In addition, customers might request changes to orders that are already in production, making it difficult to make quick decisions based on a complete, real time overview of possible options and associated consequences. Secondly, products are becoming more and more sophisticated and intelligent, with service and value added elements embedded in the products themselves. Further, with an increasing level of customisation, and demands for high quality, low costs, short and precise delivery times, and high flexibility, the management of the operations in the networks represents a significant challenge.

Other examples of E-Business Impacts on Supply Chain Integration and Business Processes are seen in Table 1.

One would imagine that the main improvements from implementing information exchange solutions would be a reduction of resources to enter/update information in the ERP-system, but experiences from Electrolux in Australia shows that even more time is released from (not) answering all kind of order related questions on the phone, not only in the administration but also in the production.

MANUFACTURING PLANNING AND CONTROL

The essential task of *manufacturing planning and control* is to efficiently manage the flow of materi-

Table 1. Examples of E-Business Impacts on Supply Chain Integration and Business Processes (Lee & Whang, 2001)

	Business Processes			
Dimensions of SC Integration	*Procurement*	*Order Fulfilment*	*Product Design*	*Post-Sales Support*
Information Integration	Supplier information sharing	Information sharing across the supply chain	Design data sharing, product change plan sharing	Customer usage data linkages
Planning Synchro-nisation	Co-ordinated replenishment	Collaborative planning and co-ordination, demand and supply management	Synchronised new product introduction and rollover plans	Service supply chain planning co-ordination
Workflow Co-ordination	Paperless procurement, auctions, auto replenishment, auto payment	Workflow automation with contract manu-facturers or logistics providers, replenishment services.	Product change management automation, collaborative design	Auto replenishment of consumables
New Business Models	Market exchanges, auctions, secondary markets	Click-and-mortar models, supply chain restructuring, market intelligence & demand management	Mass customisation, new service offerings	Remote sensing & diagnosis, auto-test, downloadable upgrades
Monitoring and Measurement	Contract agreement compliance monitoring	Logistics tracking, order monitoring	Project monitoring	Performance measurement and tracking

als, the utilisation of people and equipment and to respond to customer requirements by utilising the capacity of suppliers and internal resources to meet customer demand (Vollmann et al., 2005). *Planning* across non-hierarchical networks is a complicated and complex task due to conflicting objectives and continuously changing demand. In such a case, the planning and scheduling issues will have to face the challenge of coordinating systems that do not fully share all relevant information between companies (Alvarez, 2007).

Distributed planning in non-hierarchical networks has to be performed in a setting where relevant information from several IT systems are integrated and up to date, and can be accessed in real time from anywhere in the network. This information visibility depends on the exchange of critical data required for the efficient management and control of the flow of products, services and related information between members in the network (Handsfield and Nichols, 2002). Each node should ideally be able to see the real time situation in the network, downstream as well as upstream, from boardroom to shop floor, enabled by *automatic data acquisition*. Although *information visibility* and *system integration* are regarded as keys for enabling collaboration, very few networks have successfully achieved this (Quinn, 2003).

The dominating techniques and systems for planning and control (e.g. Enterprise Resource Planning - ERP, Manufacturing Execution Systems - MES) do not support the need for a network perspective. They have a single-company focus and mainly support centralized production and planning and control (Alvarez, 2007). This makes network planning difficult and inefficient. This amplifies the need for supply chains to rely on flexible, adaptable and responsive planning and *decision support tools* across the entire network (Özbayrak, 2006).

ADVANCED PLANNING AND SCHEDULING

Advanced Planning and Scheduling (APS) systems introduced the benefits of constraint-based planning and optimization to the business world and enabled companies to optimize plans according to financial and other strategic objectives. APS seeks to find feasible, near optimal plans unlike traditional Enterprise Resource Planning (ERP) systems (Stadtler & Kilger, 2005). APS takes into account constraints at enterprise level as well as at plant level. Materials and capacity issues are considered simultaneously, and manufacturing, distribution, and transportation issues are integrated. Many ERP and APS systems make it possible to include suppliers and customers in the planning procedure and thereby optimize a whole supply chain on a real-time basis (Wortmann, 1998; Vollmann et al., 2004; Kennerly & Neely, 2001).

Therefore, APS does not substitute but supplement existing ERP systems. The ERP system handles the basic activities and transactions such as customer orders and accounting whereas the APS system focus on the operational activities related to decision-making, planning, scheduling and control of a supply chain and related management activities, which are not explicitly well covered in ERP systems. APS has the capability to simulate different scenarios for decision support, to plan and to schedule on-line as well as off-line. Planning and scheduling is by nature a proactive process, which can initiate an event in another area of the business or at partners based on the workflow control.

APS is the most comprehensive system for supply chain planning and control today. Although APS aims at automating and computerizing the planning processes by use of simulation and optimization, the decision-making is still made by planners with insight in the particular supply chain and know how on the system constraints but likewise important: a feeling for feasibility of created plans. Thus, APS aim to bridge the gap between the supply chain complexity and the day-to-day operative decisions. This requires, however, that planners are able to model and setup decision rules for the planning and optimization.

According to Petroni (2002) there are several problems involved in using planning software such as high complexity, lack of training and knowledge among managers and personnel, low-data accuracy, and lack of support from the software vendor. Early adopters of APS report significant reductions in cycle time, resource and inventory load (OM Control, 2001; Schell, 2002; Hess, 2002). A study by Funk (2001) however, showed that only 20% of the APS installations investigated were successful (based on a threshold of achieving 70% of projected gain to become a success).

In spite of the supply chain functionality, most APS implementations are limited to a single organization or a single manufacturing site. A few cases can be found in literature on successful supply chain implementations. Gupta et al. (2002), for example, describe a decision support system which helps Pfizer to plan their distribution network. Another study from the Vita Group show an increase in delivery accuracy from 79 up to 99% while reducing lead time from 5-7 days to zero and reducing the planning resources by 30% (Works Management, 2008).

THE WORK BENCH CONCEPT

As pointed out by the US National Institute of Standards and Technology, the 20th century, competitive advantage was defined by the production and labour capabilities of individual original equipment manufacturers (OEMs). In the mid 90s, OEMs sought to reduce those costs by distributing those capabilities across a global supply chain. Competitive advantage is now seen as the combined capabilities of the suppliers that make up the OEMs' supply chain. Therefore, the only way to improve competitive advantage is to improve those combined capabilities through better integration – that is information exchange across the supply chain. The challenge is to develop and demonstrate an open, standards-based, testing and integration infrastructure that enables the automated exchange of information across the supply chain. This infrastructure will provide the foundation for new types of collaboration and management and it will help propel both OEMs and SMEs to a better competitive position in the global marketplace.

The Work Bench Concept is an alternative approach to enable such an infrastructure supporting supply chain planning where decentralised planning decisions are made with regards to individual company operations – but with a network perspective. The main functionality is to operate and manage existing SMEs manufacturing networks. The work bench comprises components for integration of information systems, visualisation of the planning and production situation, communication to enable cooperative decision making under uncertainty, optimisation of plans and simulation of the decisions, network diagnostics and performance monitoring among others. This involves a number of challenges such as providing members access to network-wide real time information (Wang and Wei, 2007), enable visualisation of the available information (Boyson et al., 2003), secure the interaction between advanced ICT based decision support tools and human decision making (Barthélemy et al., 2002), and creating a coordinated and collaborative environment (Deek et al., 2003) for planning and decision making. The state of the art, major challenges and identified requirements are listed in Table 2.

In the following section a supply chain planning approach to solve requirement 2 (Techniques for solving distributed and dynamic planning) is presented.

Table 2. State of the art, challenges and requirements for future supply chain planning solutions

State of the art	• Collaborative practices and methodologies have been developed for coordination between long term supply chain partners. Examples include Vendor Managed Inventory and Collaborative Planning, Forecasting and Replenishment (Holweg et al., 2005). • Predominant manufacturing planning and control techniques are based on centralized production control, with infinite capacity, constant lead time and fixed routings in the single company (Alvarez, 2007). • Decision support tools for network planning and execution are based on optimisation and simulation such as Advanced Planning and Scheduling, Strategic Network Planning etc. (Turban, 2007)
Challenges and conflicts	Existing collaborative practices and methodologies: • Are not supported by low cost ICT-systems • Have a single-company perspective, fail to provide collaborative planning for the network and fail to include the non-hierarchical perspective • Are restricted to centralized planning and control, and lack flexibility and real-time information. • Are out of reach for most SMEs due to the centralized control, high complexity and high cost • Assume centralized control in hierarchical networks and consensus between stakeholders Furthermore, the tools does not: • Address uncertainty and risk adequately • Support a more diversified classification of tasks to enable improved planning results based on a more holistic view. • Enable effective human interaction in decision making • Include presentation and real time information from network • Support an "individual" real-time plan generation
Requirements	1. Planning and control techniques and principles with the capability to handle conflicting objectives, dynamic networks and distributed decision making 2. Techniques for solving distributed and dynamic planning. 3. Full integration with dashboard solutions, allowing real-time decision support with advanced what-if capabilities and graphical, interactive planning boards 4. Simple ICT-supported collaborative methodologies for non-hierarchical network collaboration which: • Consider new manufacturing planning and control techniques for non-hierarchical networks • Explicitly address operational (and selected tactical and strategic issues) as well as operational uncertainties and risk, and cope with the dynamicity and complexity of manufacturing systems • Automate routine tasks, leaving human schedulers to concentrate on exceptions handling

A NEW SUPPLY CHAIN PLANNING APPROACH

Whilst the planning and optimisation procedure used in Advanced Planning and Scheduling systems is the current state-of-the-art for a single company or a hierarchical supply chain/network, planning among individual companies belonging to several non-hierarchical networks require other solutions. The reason for this is quite simple: an APS system need to have a joint goal and clear constrains for the planning engine to work, and this is not obtainable when dealing with several non-hierarchical networks. Furthermore, the hierarchical demand planning approach is based on the premise of aggregation and disaggregation which is reflected in the master demand schedule as well as the structures for capacity and material, the bill-of-resources and bill-of-materials structure. Today's APS systems planning and control capability depends on the ability to aggregate and disaggregate plans and the data they are based on. This makes the structure of the planning foundation critical, unfortunately a simple linear disaggregation of information is not always possible due to the nature of the planning foundation and the multiple usage purposes of the information contained herein.

We therefore present a relative simple solution based on individual ERP or APS plans in the respective companies followed by improvements in between partners. It is assumed that requests from customers (customer orders) have been accepted as far as possible by the ERP/APS system, possibly based on customer priorities (se below) in case of resource or materials shortage.

However, such a plan is normally based on standard customer lead-times to allow the individual ERP-plans to create realistic plans at the first attempt. The option of optimising the joint plan is therefore primarily based on the difference in between standard customer lead-time and actual lead-times in the company.

Solving the challenges related to collaboration in Non-hierarchical network is not new. A concept of Extended Value Chain Management (EVCM) is introduced by Görlitz (2002) as an add-on to the existing ERP-system. The idea is that the EVCM acts as a broker in between a customer and suitable suppliers in the market place. For the broker to identify one or more suppliers each SMEs needs to register on the market place. This solution, however, does not support the option of adjusting plans in between a customer and a supplier to enable a mutual (optimal) solution.

The outmost simple solution is therefore to allow partners to (manually) view the (production) plan and based on this utilise/request the available “free space”. A more far-reaching solution would be to allow customers to also move low priority jobs in order to create extra “free space” and thereby make way for more orders. This solution requires that jobs and customers are segmented and assigned a given priority such as illustrated in Table 3.

Assigning customer priority is a manual task. Job priorities need to be assigned automatically to allow the system to work without consuming extra resources. High priority jobs are typically related to bottleneck resources or customer orders with little or no slack. Low priority jobs could either be related to customer orders with slack or it could be stock orders. Cancellation or reduction of stock orders (jobs) might lead to higher costs depending on the details (in case of high setup costs the ERP-system might generate a stock order in connection with a customer order to reduce the total costs).

The planning approach could work in a more or less manual version where the customer-initiated changes are updated manually, but a more integrated approach would be more beneficial for all parties but also more costly. In both cases the production plan needs to be generated individually for each partner, allowing him to only view details of own orders/jobs whereas other orders/jobs only appear as a coloured (red/yellow/green) anonymous job. Further, no matter how information is changed in the system, the partners in the network likely request to keep track of the changes made. VMI vendors have already invented solutions for this and could be used as an initial source for ideas.

A strong combination of EVCM and the concept described above might lead to a more automated Collaborative Demand and Supply Planning (CDSP) model. This requires that standard interfaces are developed in between the ERP-systems and the CDSP, at least in selected areas such as items and order specifications. The process of exchanging data involves three steps:

- electronic data interchange through a message broker

Table 3. Job and customer priority rules

Jobs: Customers:	High priority (red)	Medium priority (yellow)	Low priority (green)
High priority	Orders are unlikely to be moved (customer requests change)	Orders might be moved (customer requests change)	Orders can be moved (customer initiate change)
Medium priority	Orders can’t be moved	Orders are unlikely to be moved (customer requests change)	Orders might be moved (customer requests change)
Low priority	No access to view plans		

- matching relations (e.g. dealing with different item id's, date-formats, units, timezones etc)
- process integration (in case of systems integration)

One might ask why any company should allow customers to make changes to current production plans and the answer is quite simple: to be attractive and flexible and thereby increase their competitiveness for high- and medium-priority customers. A simple comparison with our own diary would immediately create an understanding of the potential of segmenting and colouring our appointments - avoiding that we are left out of important meetings/activities because of minor/moveable arrangements.

CONCLUSION

In the highly competitive market that most companies find themselves in it is vital to organise the manufacturing and supply chain operations in the best possible way. The first step in this development process is to organise the internal operations in an appropriate and fruitful way, but for most companies there are similar improvements to be made by involving supply chain partners in the planning process. Information access and visibility, new planning and control principles, collaborative mechanisms and new information tools have to be developed in order to meet the particular challenges of SMEs in non-hierarchical networks. The paper discusses the current state of the art applications and outlines a new approach titled "Work Bench" to further improve the planning with special focus on non-hierarchical networks. Especially a simple approach to align production plans is presented.

ACKNOWLEDGMENT

We would like to acknowledge the other partners (SINTEF, Frauenhofer, IMI, ISOIN and UNIBG) of the PLANET FP7 proposal for valuable contribution.

REFERENCES

Alvarez, E. (2007). Multi-plant production scheduling in SMEs. *Robotics and Computer-integrated Manufacturing*, *23*, 608–613. doi:10.1016/j.rcim.2007.02.006

Ayers, J. B. (2006). *Handbook of supply chain management* (2nd ed.). New York: Auerbach Publications.

Brown, G., Keegan, J., Vigus, B., & Wood, K. (2001). The Kellogg company optimizes production, inventory and distribution. *Interfaces*, *31*(6), 1–15.

IOM Control. (2001). ERP survey. *Institute Operation Management, 27*(8).

Danese, P., Romano, P., & Vinelli, A. (2004). Managing business processes across supply networks: the role of coordination mechanisms. *Journal of Purchasing and Supply Management*, *10*, 165–177. doi:10.1016/j.pursup.2004.11.002

Funk, G. (2001). Enterprise integration: join the successful 20%. *Hydrocarbon Processing, 80*(4).

Görlitz, O., Neubert, R., Teich, T., & Benn, W. (2002). Extended Value Chain Management on Electronic Marketplaces. *International Journal of e-Business Strategy Management, 3*(3).

Gupta, V., Peter, E., Miller, T., & Blyden, K. (2002). Implementing a distribution-network decision-support system at Pfizer/Warner-Lambert. *Interfaces*, *32*(4), 28–45. doi:10.1287/inte.32.4.28.54

Handsfield, R. B., & Nichols, E. L. (2002). *Supply Chain Redesign: Transforming Supply Chains into Integrated Value Systems*. Upper Saddle River, NJ: Prentice Hall.

Harland, C. M., Lamming, R. C., & Zheng, J. J. (2001). A taxonomy of supply networks. *Journal of Supply Chain Management*, 37.

Hess, E. (2002 May). Make advanced planning & scheduling work for your company. *APS*.

Holweg, M., Disney, S., Holmström, J., & Småros, J. (2005). Supply Chain Collaboration: Making Sense of the Strategy Continuum. *European Management Journal*, *23*, 170–181. doi:10.1016/j.emj.2005.02.008

Jagdev, H. S., & Thoben, K. D. (2001). Anatomy of enterprise collaborations. *Production Planning and Control*, *12*, 437–451. doi:10.1080/09537280110042675

Kennerly, M., & Neely, A (2001 February). Enterprise resource planning: analysing the impact. *Integrated Manufacturing System*, 103-113.

Lee, H. L., & Whang, S. (2001). E-Business and Supply Chain Integration. *Stanford Global Supply Chain Management Forum*. Retrieved from http://www.stanford.edu/group/scforum/Welcome/EB_SCI.pdf

Özbayrak, M., Papadopouloub, T., & Samarasb, E. (2006). A flexible and adaptable planning and control system for an MTO supply chain system. *Robotics and Computer-integrated Manufacturing*, 22.

Petroni, A. (2002). Critical factors of MRP implementation in small and medium-sized firms. *International Journal of Operations & Production Management*, *22*(3), 329–348. doi:10.1108/01443570210417623

Quinn, F. (2003). The Elusive Goal of Integration. *Supply Chain Management Review*, 7.

Sahin, F., & Robinson, E. P. (2002). Flow co-ordination and information sharing in supply chains: review, implications and directions for future research. *Decision Sciences*, *33*(4), 1–32. doi:10.1111/j.1540-5915.2002.tb01654.x

Schell, D. (2002). Overcoming intricacies of ERP system implementation. *Integrated Solutions*, 2.

Stadtler, H., & Kilger, C. (2005). *Supply Chain Management and Advanced Planning – Concepts, Models, Software and Case Studies* (3rd ed.). Berlin: Springer.

Vollmann, T. E., Berry, W. L., Whybark, D. C., & Jacobs, R. F. (2005). *Manufacturing Planning and Control for Supply Chain Management*. Boston: McGraw-Hill.

Works Management. (2008). *IT in practice Planning & scheduling, Growing in stature*. Retrieved April 2009, from http://www.worksmanagement.co.uk/articles/12830/Growing-in-stature.pdf

Wortmann, J. C. (1998). Evolution of ERP systems. In Bititci, U. S., & Carrie, A. S. (Eds.), *Strategic Management of the manufacturing value chain* (pp. 11–23). Boston, MA: Kluwer Academic.

This work was previously published in Enterprise Information Systems Design, Implementation and Management: Organizational Applications, edited by Maria Manuela Cruz-Cunha and Joao Varajao, pp. 496-504, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 9
Instructional Design of an Advanced Interactive Discovery Environment:
Exploring Team Communication and Technology Use in Virtual Collaborative Engineering Problem Solving

YiYan Wu
Syracuse University, USA

Tiffany A. Koszalka
Syracuse University, USA

ABSTRACT

This chapter examines the instructional design of, and reports on research conducted within, a multiuser virtual environment created for a distributed Collaborative Engineering Design (CED) course. The course's Advanced Interactive Discovery Environment (AIDE) provided a variety of synchronous online tools and communication devices to support SameTime virtual team collaboration and problem-solving within the course. The research helped to unpack (1) which tools team members engaged with during collaborative learning activities, (2) how and why they used or did not use provided online features to support their individual learning and enhance team productivity, collaboration, and communication, and (3) how team members communicated socially. The research also describes how different team social communication patterns may be related to the patterns of team technology use. Relevant theoretical frameworks including social learning, media stickiness, cognitive imprinting, and recommendations on how different tools can be effectively integrated into multiuser virtual environments to facilitate learning are discussed.

DOI: 10.4018/978-1-4666-1945-6.ch009

INTRODUCTION

Dieterle and Clarke (2005) advocated that the best learning environments for students are those that are authentic, situated, and distributed across internal and external sources. With advances in computational technologies and network connectivity, Multiuser Virtual Learning Environments (MVEs) and Computer-Supported Collaborative Learning (CSCL) have become more integrated into instructional environments to provide learners with the tools to communicate synchronously in virtual ways. Simultaneously, authentic problem-solving experiences incorporating simulations and sharing applications like electronic whiteboards also began to be integrated into these new multiuser virtual learning environments to create a real-world situated learning experience. Educational environments were thus beginning to be designed to engage learners in the complex distributed collaborative activities experienced by practicing engineers and scientists (Shuman et al., 2002). Such environments can provide learners with dynamic, interactive multiuser virtual spaces in which they are able to engage in rich hands-on learning experiences, apply new content knowledge to legitimate problems, collaborate in problem solving, and use the emerging technologies of their practice.

Geer and Barnes (2006) suggested that with the rapid development of emerging technologies and virtual environments educators face a predicament of deciding which technologies are best suited to support expected learning. Given a lack of knowledge of emerging technologies and lack of pedagogical guidance about integrating technologies for collaboration and communication, educators are often left with mounting dilemmas and confusion about which technology-based resources are most effective for given pedagogies and learning expectations. In addition, the lack of understanding about how students learn in complex distributed collaborative activities and how their cognitive processes operate within synchronous virtual learning environments inhibits the design and implementation of effective and appropriate CSCL environments. As Pellegrino (2006, p.3) stated:

…most current approaches to curriculum, instruction, and assessment are based on theories and models that have not kept pace with modern knowledge of how people learn. They have been designed on the basis of implicit and highly limited conceptions of learning. Those conceptions tend to be fragmented, outdated, and poorly delineated for domains of subject matter knowledge.

In addition, little research has been provided that clarifies the role of communication in multiuser technology use, learning, and problem solving within synchronous computer-mediated learning environments. Limited research has focused on observations of team members becoming more task-oriented and developing clearer role expectations among themselves in online learning communities, with little focus on content or process learning (Jonassen & Kwon, 2001). Numerous researchers who study such multiuser virtual environments have found no conclusive evidence that these environments work, and when and most importantly why they do or do not work (Kirschner et al., 2004). Therefore, additional theories and confirming research are needed to better inform the instructional design (ID) of multiuser virtual environments that facilitate learning of subject matter and appropriately incorporate emerging technologies in pedagogically sound ways.

THEORETICAL FOUNDATON: SOCIAL COMMUNICATION, COGNITIVE IMPRINTING, MEDIA STICKINESS AND THEIR IMPLICATIONS TO ID

Social Communication to Learning in Multiuser Virtual Environments

From social constructivist perspectives, learning is observed to be a social process by which knowledge is generated through a network of collaborative interactions and is distributed among humans and tools that interact (Lowyck & Poysa, 2001). Collaborative learning theories suggest that effective learning requires learners with various types and levels of knowledge to work together toward a common knowledge development goal (Gokhale, 1995). Peers are key players in the human learning process as working together can be beneficial in helping each gain better understanding by 'tossing around' diverse ideas, considering the strengths and weaknesses of each idea, and revising their current knowledge and beliefs in light of new evidence and insights from others (Ormrod, 2008). During social interactions each learner constructs his or her own knowledge and skills and contributes to team member learning (Panitz, 1996).

A learner's cognitive growth can be promoted by peer interactions. Such interactions also tend to stimulate the learner's affect or feelings and motivations. For example, many learners find interactive collaborative learning sessions highly motivating, in part, because such environments address their social needs while they are studying subject matter (Ormrod, 2008). As Chen (1994) observed, students who felt uncomfortable in an educational communication environment avoided social interaction, were less argumentative, and were less willing to advocate their position or challenge other students on controversial issues. Fåhræus (1999) found communication pattern differences among productive and unproductive groups, specifying that the content of productive groups' messages contained higher frequencies of greetings, personal descriptions of participants' professional, social, technological context and questions, and feedback among group members.

Rourke and Anderson (2002) warned that although a moderate amount of social communication created a productive climate for learning, social communication that overtook critical discourse exasperated some learners. Learners welcomed interpersonal interaction during online collaborations. However, if the collaboration became too social and moved away from the task-orientation, learners became very dissatisfied. Thus, social-collaborative learning activities, focusing on learning tasks with a non-disruptive level of informal social communications, are impacting the learners' affective domain (e.g., feelings, emotions) and support cognitive development (learning).

Virtual multiuser environments involve complicated social patterns and communication strategies that often require multiple synchronous communication channels (e.g., voice, video, document sharing, whiteboard features) to engage learners in collaborative tasks and learning. Rourke and Anderson (2002) and others (e.g. CSCL researchers) suggested that further investigation should be made into the nature of these relationships between social communication and learning, especially in synchronous multiuser educational environments where little research has been shared.

Several factors have been identified in prior research that mediate a virtual team's communication, interaction, and its learning process. These factors include learners characteristics, prior content knowledge, cultural differences, provision of a CMC coach, and content of the course (Geer & Barnes, 2007; Hron & Friedrich, 2003; Prinsen, Volman, & Terwel, 2007; Ross 1996; Solimeno et al., 2008). To disclose and understand team communication within virtual multiuser learning environments, this study examined four social communication variables identified as important to synchronous CSCL team communications including the: (1) level of team member participation (Kapur & Kinzer, 2007; Prinsen, Volman, & Terwel, 2007; Ross, 1996); (2) amount of team socialization – time in off-task activities (Rourke & Anderson, 2002); (3) amount of in-team time on-task (Jonassen & Kwon, 2001); and (4) team communication [strategy] differences (Cho et al., 2007; Nussbaum et al., 2004). This data was considered with the team members' overall evaluation of their synchronous collaborative communication experiences and studied in light of the types of technologies provided during individual and collaborative learning activities.

Cognitive Imprinting Theory

Cognitive imprinting (Geer, 2005) accounts for a tendency of multiuser virtual learning groups to repeat early patterns of communication, which can indicate their cognition levels, as they collaboratively solve a problem as a team. Learner initial communication patterns were shown to be powerful in determining subsequent interactive behaviors in synchronous multiuser learning communities. Kapur and Kinzer (2007) found that inequities in team member participation patterns exhibited a high sensitivity to initial exchange and patterns tended to get locked-in early in the discussion, ultimately lowering the quality of discussion and, in turn, the group performance. Furthermore, Geer and Barnes (2007) suggested that the notion of cognitive imprinting can be taken as a means of characterizing the serially consistent cognitive behavior of the students within technology-facilitated learning environments, and further may reflect desired learning outcomes.

Technology Implementation: Media Stickiness Theory

Huysman et al. (2003) suggested the concept of media stickiness to explain the tendency of individuals or groups to maintain early choices of technologies despite the introduction of more useful and value-added technologies. Although distributed virtual teams exhibited distinctly different patterns and preferences for communication, collaboration, and learning technology choices, their choices were generally found to be made early in project work and remained the preferred practice throughout the project.

Both cognitive imprinting and media stickiness imply that team communication and technology adoption patterns rely on student initial cognitive behaviors. Once these cognitive behaviors became habitual the chance of changing patterns was limited. This may in turn constrain student learning of new concepts, skills, and knowledge, especially in creativity and higher order cognition development. Both theories seem to conflict with Rogers' theory of innovation adoption. Rogers (1983) suggested that adopting innovations, like new technology tools, concepts, or skills, is a process of passing from an initial stage of knowledge of an innovation, into a stage of forming an attitude toward the innovation, into a stage of making a decision to adopt or reject the innovation, into a stage of implementing (using) the new innovation, and into a final stage of making a decision to adopt [or not] the innovation.

To support adoption of synchronous multiuser collaboration tools, for example, training should help students develop knowledge and practice using the new technologies. Training should also provide instruction in technology-based collaborative communication skills (e.g., team management strategies) while encouraging students to experiment with the new communication technologies within the virtual environment. As Grabowski and Koszalka (2002) suggested, continuously engaging students with new technologies and repeating such activities helps students form knowledge and attitudes toward innovations that bring them value and aid in adoption.

However, knowledge and practice are not always sufficient in supporting the adoption process. As to technology adoption, it is still not clearly understood whether a lack of knowledge and attitude toward a new technology or a failure in the implementation (use) of a new technology is to blame for lack of adoption. More specifically, no empirical data was found that indicates sticking to an initially chosen technology is a failure in knowledge of a new technology or were there data found that indicate in which stage of the innovation-decision process it is most likely where technology adoption failures may occur. In addition, the concept of habitual inertia (inaction), implied in both the cognitive imprinting and media stickiness theories, is valuable in describing contexts where a constant and repetitive performance is expected to be developed. Therefore, examination of students' team learning processes and understanding of instructional and learning needs will help in understanding when learning inertia may occur, whether it needs to be maintained or stopped, and what instructional interventions may be appropriately designed and implemented to support technology adoption choices.

RESEARCH CONTEXT

A collaborative engineering design (CED) course was created to engage distributed teams of engineering students in a multiuser, blended synchronous and asynchronous, virtual environment to learn about and solve authentic engineering design problems. The CED incorporated a virtual multiuser work environment, the Advanced Interactive Discovery Environment (AIDE), to engage learners with a variety of tools that supported synchronous distributed, interdisciplinary, collaborative engineering design tasks. This environment allowed multiple learners entry into this virtual space to participate synchronously in live lectures and discussions to learn about and apply engineering concepts. Learners were able to employ a variety of communication and analysis tools in a virtual large class setting or in small team activities. While inside the AIDE, learners were able to share ideas and explore solutions orally and visually through audio and video conferencing, shared writing and drawing spaces, and data analysis application sharing. Dietele and Clarke (2005) identified that "…multiuser virtual environments enable multiple simultaneous participants to be able to access virtual contexts, … communicate with other participants, … and take part in experiences incorporating modeling and mentoring about problems similar to those in real world context". The CED AIDE SameTime environment created a significant, interactive, virtual collaboration learning experience by offering innovative instructional ways to students for knowledge sharing and decision-making tasks.

CED COURSE DESIGN

The Study Context, Instructional Design Foundations, and Multiuser Virtual Environments

Instructors from two universities collaboratively taught the semester-long CED course directly to the students from their home institute and synchronously through the AIDE at a distance for those at the partnering institute. Participating students from both universities simultaneously attended course lectures either in-person at their home institute or through multiuser virtual environment synchronous tools depending on which of the two professors was responsible for the session content. Students were assigned to distributed design teams with 50% from the local and 50% from the distance university. Each received instruction in the foundational engineering content and necessary technology skills to participate in the course activities. For several weeks however, the students were split into one of two engineering content learning tracks. Thus students, for the sake of a culminating activity, had different engineering expertise from which to collaborate on a resolution for a given engineering design problem, which required students to create a preliminary design of a thermo-structural system for a specific location on a hypothetical second-generation reusable launch vehicle (RLV) for NASA space missions. Each distributed team therefore had a mix of students from both universities and each of the two engineering content tracks. These distributed engineering design teams thus had to bring together different types of engineering knowledge to collaboratively solve a design problem in an authentic engineering working environment, that is, as distributed interdisciplinary teams working to solve complex engineering problems. The course instructors acted as project managers and mentors to help the teams be successful at their tasks.

The Advanced Interactive Discovery Environment (AIDE) for Engineering Education was developed in partnership between the faculties of the two universities. The AIDE integrated and advanced the best features of virtual, collaborative engineering environments, state-of-the-art simulation tools, and advanced learning management systems (Davidson et al., 2002). Pedagogical approaches, including generative, problem-based, and collaborative instructional strategies formed the structure of the course to fully engaged learners with the AIDE to learn content and learn how to resolve engineering design problems.

The course was also designed to promote skills development in using leading-edge technologies, working on distributed teams, writing project reports, and giving oral presentations. Teamwork was a critical aspect of the course design thus several team building exercises were presented. Early in the semester *best practices* labs were conducted to help students build productive teams using the collaboration technologies available in the AIDE.

Course Technology

The course infrastructure was built with four main components: (1) distance learning classrooms (DLC) equipped with video conferencing system for the synchronous full class instruction – See Figure 1, (2) IP-based web-conferencing to satisfy the needs of multiple simultaneous partial-class events and synchronous team collaboration, (3) asynchronous online information management and communication system provided a platform to document course materials, exchange files, and store engineering discipline-related resources, and (4) a physical design studio and individual tablet-pc provided a suite of technology tools for the students. See Table 1.

The provided multiuser synchronous communication features (SameTime [ST]) included video and audio, interactive shared whiteboard, application sharing, chats, instant messaging, document posting, and presence-awareness. The DST lectures and learners' team meetings were conducted through SameTime. The asynchronous AIDE services (e.g., course announcement, team work spaces, student drop-boxes, threaded discussion board, and team project management tools) also provided supports for team communication and learning activities. See Figure 2.

RESEARCH QUESTIONS

An initial data analyses suggested the course design fostered engineering content learning, CED problem-solving, collaborative technologies

Figure 1. Distance learning classroom. (a) Student view. (b) Faculty view

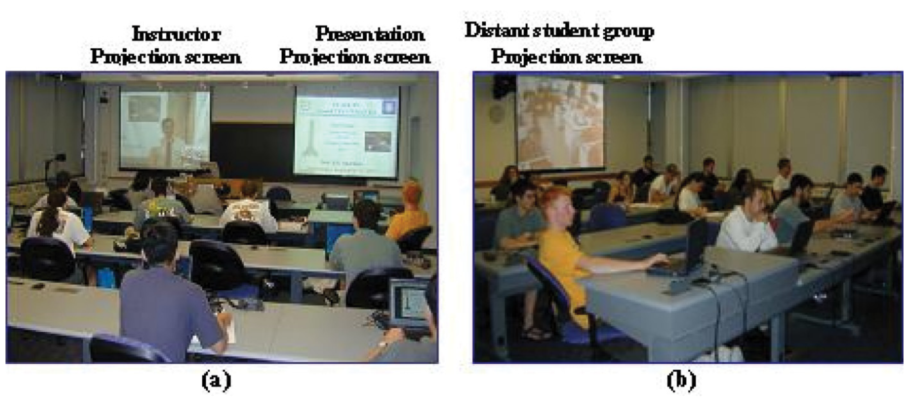
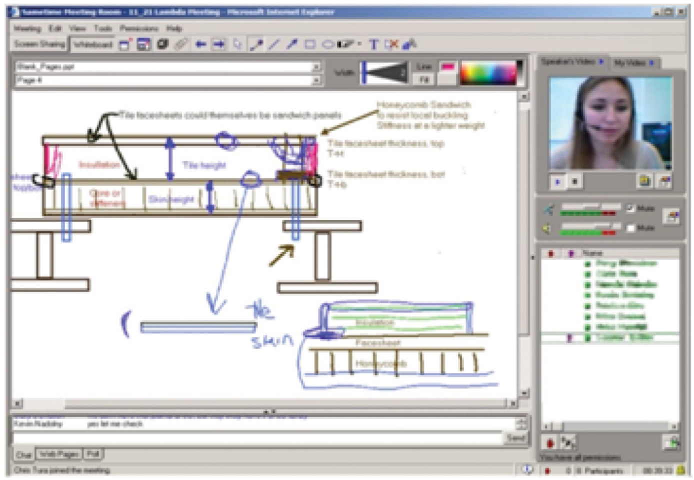


Table 1. Summary of provided classroom and virtual tools

Location	Tools	Description / Key features
Distance Learning Classrooms	• Video conferencing equipment,	Facilitate delivery of instruction and 2-way communication and interaction
	• Projection equipment and electronic white board (facilitate local and distance education)	Facilitate 2-way communication and interaction
	• SmartBoard	Facilitate presentation, document, etc. sharing, communication, and interaction
Design Studio	• Physical lab (at each site) with computer stations	Engineering and analysis software loaded stations to facilitate computation and design activities; supports communication
Course hardware	• Tablet PC for individual students	Provide 24/7 access to AIDE ST for all students
Virtual (AIDE ST)	***Synchronous communications***	
	• Video / audio	Real time (taped/stored) video conferencing
	• Interactive shared whiteboard	Collaborative space where teams can draw, type, color, highlight, etc., all see and edit together
	• Screen sharing	Ability to show all team members an application or document, one person demonstrates and edits
	• Meeting Room Chat	Text-based communication, all can add and see comments posted
	• Polling	Ability for team members to vote in response to posted questions during meetings. question must be posted before a ST meeting session begins
	• Send Web Page	Ability to open and share a Web Page, one person implements and others can see
	• Hand raising	Ability for any team member to take control of talking (audio sharing) during ST session
	Asynchronous Communications	
	• Course announcements	Information, notices, reminders posted by instructors
	• Team work storage spaces	Storage space open only to specified team members
	• Student drop boxes	For submission of each student's assignments
	• Threaded discussion board	Traditional discussion board
	• Team project management tools	Templates to help teams management their progress and report to course managers

knowledge, and development of collaboration skills (Koszalka & Wu, 2008). However, two significant instructional design themes emerged from this initial analysis: designing effective uses of technology resources during collaborative activities and supporting team collaboration strategies. Learners reported that they were not aware of the value some of the provided AIDE tools, how to use them effectively in their collaborative work, or even about the existence of particular tools until late of the semester (Koszalka & Wu, 2008). This feedback appears to support Kirschner et al. (2004) conjectured that learners sometimes participated in online environments very enthusiastically, but often used the tools in a perfunctory way. Learners also tended to display different understandings of team participation and often reported that their team work suffered because of a lack of a team management planning and team structure in the virtual environment. It was not known whether the issue was in the design of the learning environment, the types of technology tools and resources provided, training on the technology tools, or instruction in team

Figure 2. Sample of Web-based synchronous meeting and discussion screen. Participants can see the current speaker, share applications, use chat, and markup whiteboard during the discussion



collaboration and the use of the technology tools during collaborative design activities.

The instructional goals of multiuser virtual environments were to equip learners with cognitive skills through which they learn how to be successful lifelong learners and problem-solvers when they are situated in different contexts facing challenges from new technologies, new people, and new problems. Hence, learners unable to use technologies or to effectively learn in a virtual multiuser CED environment may not yet have obtained the skills and knowledge to be successful in such complex and authentic environments. They may also lack the collaboration skills necessary to participate effectively in collaborative problem-solving activities. Based on these conjectures, the four research questions for this study included:

- Which AIDE SameTime (synchronous multiuser) web-conferencing technology tools did the teams use most often during the CED project? How and why do they use these tools?
- Are there different patterns between the two teams in learning and / or adopting of technologies during the team's AIDE SameTime synchronous web conferencing meetings?
- What type of technology-based communication patterns did each team display?
- Was there a relationship between team communication pattern and team use of AIDE SameTime (synchronous multiuser) technologies during meetings?

METHODS AND DATA COLLECTION

This was a mixed methods exploration of a virtual multiuser collaborative engineering design course with a focus on how learners engaged with the provided synchronous communication tools. A stratified random sample of the synchronous multiuser AIDE SameTime team recorded meetings was identified to ensure that data represented sessions held at the beginning, middle, and end of the course, representing all four units in the course.

The sampled videos were reviewed using a structured video review protocol. A SameTime Video Review Sheet (protocol) was composed of three parts developed from a combination of existing data analysis and demographic protocols and a synthesis of the literature and similar studies on social learning activities (Geer & Barnes, 2007; Huysman et al., 2003; Ross, 1996; Rourke & Anderson, 2002). Part I of the review protocol was a Technology / Media Use & Computer Tools Competence Table, adapted from Huysman et al.'s (2003) team media-use pattern table. Part II was a SameTime Tech-mediated Team Communication protocol created through the synthesis of the literature on variables of social collaboration that affect learning during distributed learning activities. Part III was a series of evaluative and demographic questions to provide descriptive information about the teams and their members. Each video was reviewed by two coders who used the SameTime Video Review Sheet to gather a variety of qualitative data.

DATA ANALYSIS

One-hundred and forty-two SameTime team meetings were recorded during the course, 83 were usable as data. A stratified random sampling technique was used to select ten videos, five each from two CED teams. The videos represented both teams and all time periods over the course. The ten videos represented approximately 855 minutes of team interactions. See table 2.

Three doctoral students were recruited and trained as coders. Each sampled video was reviewed by two coders. The inter-rater reliability of the coding was 89.96%. All data was imported into QDA Miner software for coding. Seventy-two sub-codes were produced under four major categories: (1) demographics, (2) technology use, (3) team communication, and (4) reflection.

RESULTS

Overall, the data suggested that the two teams had different technology use patterns and communication strategies. Specific results for each research questions are presented below.

R. Q. 1: Which AIDE SameTime (Multiuser Synchronous) Web-Conferencing Technology Tools Did the Team Use Most Often during the CED Project? How and Why Do They Use These Tools?

Major SameTime Tools Used

It was observed that audio, video, interactive whiteboard, window chat room and screen sharing were the five major SameTime multiuser tools most frequently used in the sampled synchronous meeting videos.

Interactive whiteboard: The WB was used when the team members collaborated on the team project. The most frequently used Whiteboard features included the pen, presenting, changing slides, and eraser. The two teams seemed comfortable using different editing and communicating features provided by the interactive whiteboard. They used these tools to review team documents, take notes, brainstorm, visualize their thoughts through writing or drawing, and cooperate on

Table 2. Descriptive information of analyzed team communication videos

Team	Sept. 12	Sept. 15	Sept. 21	Oct. 5	Oct. 27	Oct. 31	Nov. 5	Nov. 14	Nov. 29	Total video minutes
Alpha	1		1	1		1		1		484
Gamma		1		1	1		1		1	371
Subtotal	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**855**

scheduling and team logistics, or to merely scrawl for fun.

Window chat room: This feature was frequently used to (1) backup to audio for brainstorming, (2) express an opinion, inform, comment, send out a query, or respond without interrupting the audio, or (3) joke, give a salute or acknowledge to teammate without interrupting the audio.

Screen sharing: The AIDE SameTime Screen Sharing features were used regularly by the Alpha team. The screen sharing does not allow multi-editors. The person sharing the screen took notes on his or her screen during team discussions and the other team members ensured that all changes were recorded correctly as they read the shared screen. As was observed, uploading the screen sharing document was time consuming and running the screen sharing tool slowed the performance of other SameTime tools (e.g., audio, video). The screen sharing feature was used less frequently by the Alpha team toward the end of the semester.

Computer audio and video: The audio and video features were the most commonly used synchronous communication features of the AIDE SameTime.

R. Q. 2: Are there Different Patterns Between the Two Teams in Learning and / or Adopting Technologies during the Team's AIDE SameTime Synchronous Web Conferencing Meetings?

Team Differences in Multiuser Technology Uses

At the beginning of the semester, both teams were observed practicing different technologies and tools to a certain level within the SameTime meeting environment. The two teams presented different technology use patterns and these differences included:

Number of tools used: The Alpha team members were observed using ST tools extensively while the Gamma team members were observed using only the basic meeting features, such as presenting documents and pen tools that the interactive whiteboard supports. Among the 14 mini-features of the interactive whiteboard, the Alpha team experimented and used 13 while the Gamma team only used 7 (see Tables 3 & 4). In addition, Gamma team was rarely observed using screen-sharing and send-web-page features in the sampled SameTime videos.

Use frequency of major tools: The Alpha team used interactive whiteboard 3 times more often and the window chat room 10 times more often than the Gamma team. In addition, the Alpha team used the screen sharing application four times while the Gamma team never used it (see Figures 3 and 4).

The window chat room was the technology reported most commonly used by the students in their daily life routines outside of classes; therefore, it was presumed that students felt more comfortable using chat than other SameTime technologies. Data suggested that the Alpha team used chat very commonly throughout the entire semester while the Gamma team was observed using the chat feature less often. The chat was used as the major salutation tool by the Gamma team when someone accessed the meeting late. As the Gamma team became more familiar with the SameTime environment and the audio equipment, they generally used oral conversation, rarely using the chat or other SameTime communication tools.

Issues and difficulties: The more frequently and higher number of technology tools used during sessions the more issues or difficulties arose. For example, the Alpha team used more different tools, more often, than the Gamma team. The Alpha team also encountered 43 technology issues, 8 times more than those encountered by the Gamma team. See Table 5.

Patterns of technology use: Both the Alpha and Gamma teams were observed experimenting with the SameTime technology regularly at the

Table 3. Team Alpha tech use frequency

		Alpha Sep12	Alpha Sep21	Alpha Oct05	Alpha Oct31	Alpha Nov14	Total
	Arrow	1	0	0	0	0	1
	Changing slides	7	2	2	9	1	21
	Circle	0	0	1	0	0	1
	Colors	9	0	18	6	0	33
	Creating a New Slide	0	0	1	1	1	3
	Eraser	1	0	21	10	11	43
	Line	1	0	0	0	0	0
Interactive Whiteboard	Pen	10	2	42	49	26	129
	Presenting/Uploading	3	1	2	8	2	16
	Rectangle	2	0	7	0	0	9
	Refreshing	0	0	2	0	0	2
	Selection	0	0	1	0	0	1
	Thickness	0	0	0	0	0	0
	Typing	4	0	9	1	7	21
	Total	**38**	**5**	**106**	**103**	**48**	**300**
Screen Sharing		**0**	**3**	**0**	**1**	**0**	**4**
Send Web Page		**0**	**0**	**0**	**0**	**0**	**0**
Window Chat Room		**6**	**29**	**58**	**15**	**103**	**211**

beginning of the project. Their technology uses continuously increased until the middle of the semester. Pen and chat became the two dominant collaboration tools for the Alpha team. However, the Gamma team dramatically decreased their use of technology and seldom used the SameTime tools toward the end of the semester.

Figure 3. Team Alpha technology use pattern (four major tools used)

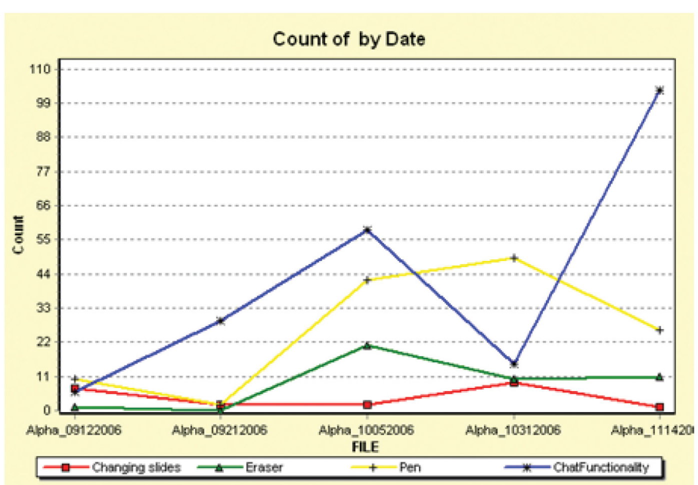


Table 4. Team Gamma tech use frequency

		Gamma Sep15	Gamma Oct05	Gamma Oct27	Gamma Nov05	Gamma Nov29	Total
Interactive Whiteboard	Arrow	0	0	0	0	0	0
	Changing slides	4	25	0	0	6	35
	Circle	0	0	0	0	0	0
	Colors	2	1	0	0	4	7
	Creating a New Slide	0	1	1	1	1	4
	Eraser	1	4	1	0	2	8
	Line	0	0	0	0	0	0
	Pen	5	38	9	9	13	74
	Presenting/Uploading	1	5	1	1	2	10
	Rectangle	0	0	0	0	0	0
	Refreshing	0	0	0	0	0	0
	Selection	0	0	0	0	0	0
	Thickness	1	1	0	0	5	7
	Typing	0	0	0	0	0	0
	Total	**0**	**75**	**12**	**11**	**0**	**98**
Screen Sharing		**0**	**0**	**0**	**0**	**0**	**0**
Send Web Page		**0**	**0**	**0**	**0**	**0**	**0**
Window Chat Room		**7**	**8**	**0**	**0**	**7**	**22**

Tool choices: Both Alpha and Gamma teams were sometimes observed using less efficient tools for a task. For example, one person in the Gamma team consistently orally described his analysis work rather than using screen sharing tool to present his work. It was also observed that when both teams had to change slides or go back-and-forth more than one slide, they kept go-

Figure 4. Team Gamma technology use pattern (four major tools used)

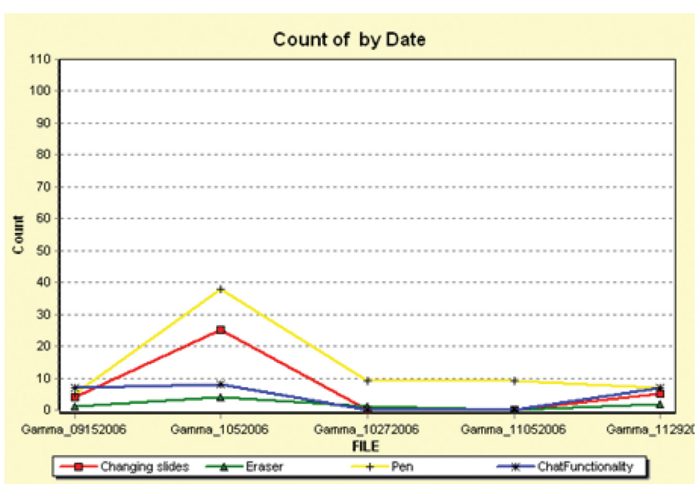


Table 5. Technology issues encountered by each team

Team	Alpha Sep12	Alpha Sep21	Alpha Oct05	Alpha Oct31	Alpha Nov14	Gamma Oct05	Gamma Nov29
Audio / Video issues	5	13	8	3	6	2	1
Other tech issues	5	2	0	0	1	2	0
Total	**43**					**5**	

ing slide-by-slide rather than using the dropdown menu to go directly to the desired slide. These choices decreased the productivity of the meeting.

Overall Evaluation of Use of Multiuser SameTime Tools

The data generated in this study confirms media stickiness existing in students' use of technologies by revealing:

Stickiness to the SameTime tools used at the early stage of the semester: In addition to the audio and video, the pen and chat were two new AIDE SameTime technologies that both teams stuck with throughout the semester. This suggests that the team members identified these tools as valuable to their collaborative work, easy to learn and use in virtual communication, and developed the skills to use them effectively; but found some of the provided tools were not efficient, or too difficult to use for their communications and problem solving activities.

Stickiness to the traditional way of using tools: Chat was promoted by the instructors and required to be used as brainstorming tool in one of the Best Practice sessions. The instructors required the teams to record their thoughts in the chat when the team brainstormed for their Ball-in-pipe assignment. However, chat was rarely observed being used as brainstorming and recording tool in the later team meetings. There was, on several occasions, one member of the Gamma team who was observed suggesting the use of the chat feature to record key ideas generated in the team discussions but his proposal was never accepted by other team members.

Traditional technologies used for communication outside of the SameTime environment: Additional data was gathered suggesting that students from both teams continued to use text messaging, cell phone, and emails most frequently outside of the ADIE to exchange documents or to inform each other of scheduling changes or other emergencies. This suggests that the student were comfortable with these everyday tools, saw value in them for communicating with their teammates, and perhaps were not as comfortable with using only the provided AIDE tools.

TEAM COMMUNICATION

R. Q. 3: What type of Technology-Based Social Communication Patterns did Each Team Display?

Team Participation

Alpha team: At the start of the semester, the moderator[1] of the meeting tended to take more initiative and be more expressive during virtual meetings. The moderator usually dominated the conversation and ensured that the team was on track. In general, the interaction among team members was well-balanced and participation of the team members was generally equal. It was often observed that two students would take one side of an issue or argument and the other two members would support a different idea or approach. After

AF[2] joined the Alpha team (mid-semester) the team dynamics were observed to change. AF appeared to have a good grasp of the content knowledge and be very focused on tasks. AF began to lead and dominate many of the discussions in a majority of the meetings.

Gamma team: In general, participation was roughly equal for members GL, BZ, and BK. MW[3] was observed to have a fair level of engagement at the semester start; however, MW's participation became less toward the semester end. BZ and GL were often observed taking over as leaders and polarizing the discussion. BK was not central to the group suggestions but his contributions were thoughtful suggesting he was listening and waiting for appropriate times to contribute. It is worth noting that when GL was not attending a meeting, the team dynamics changed slightly. BK and BZ took over as conversation leaders and BK became well conversed.

On-Task Performance

Both Alpha and Gamma teams appeared to be task-oriented and deadline driven during their synchronous virtual meetings. They each had specific tasks to accomplish in every session. The meeting usually started with technology normalization checking that all team members had the SameTime set-up and ensured proper SameTime functioning. Depending on tasks, meeting time was sometimes separated into individual work and team collaboration periods. During individual work time, each member worked on his own solution to a common problem and then the team collaborated by exchanging, discussing, and comparing individual thoughts. At the end of the meeting, the teams usually scheduled the next meeting time and discussed team logistics. Although common collaboration routines were used by both teams, on-task performance and communication patterns were different.

Alpha team: Alpha team meetings and working structures were relatively loose. Team members were observed having food or beverage during meetings, being late, stepping away in the middle of the discussion, or chatting casually. There did not appear to be a team management strategy and the team productivity was low. Feedback from one of the instructors suggested that the Alpha team had a weak plan for their design project, without specifics as to who was to complete each task and by when. The Alpha team did not have an effective scheduling process – they spent much longer in their scheduling phase than was necessary, leaving them less working time for their design project. The team spent up to 35 minutes at the meeting start trying to schedule the next session yet could not reach consensus on a time that fit everyone's schedule. After 15 minutes of content discussion, the team returned to discuss scheduling finally identifying a tentative schedule.

Content related jokes were generated during team conversations. The Alpha team members were occasionally observed teasing their team members. The jokes created by the team were entertaining and may have helped them relax. However, the jokes may also have been distracting to team productivity. It was observed that after joking around it sometimes took long periods of time to refocus on the meeting goals.

Gamma team: the Gamma team was strongly focused on project content and tasks. When the meeting started, team members briefly checked the technologies and greeted their peers. They quickly moved onto tasks. Based on the feedback from one of the instructors, the Gamma team had a reasonably good plan for their design project with dates, specific tasks, and assignments.

Socialization and Off-Task Time

Major reasons causing the team to get off-task included: waiting for other team members to login in, encountering technology problems, such as someone's voice was cut off, waiting for the screen sharing application to open, team members needing to step away, and frustration from the lack of focus during team discussions. The Alpha team was observed being off-task more than ten times. Meanwhile, it was rarely observed that the Gamma team was off task.

Team Dynamics and Communication Differences

Personality vs. Teamality: Every student appeared to react differently to a particular problem perhaps due to their personality differences. However, when the students worked as a team, every team started forming its communication characteristics, or we may name it *teamality*, regardless the diversity its team members' personalities appeared.

Alpha team examples: The five team members in the Alpha team sometimes appeared to behave differently. For instance, MA appears to like the role as moderator, often taking charge and making decisions quickly. AB appeared more detached, not participating very often, and seeking the most straightforward solutions. LS, in contrast, was observed using technology more and really wanted to tackle the solution by sticking to his own thoughts and plan. JR preferred a more pragmatic approach while settling for a feasible solution, if possible. AF joined the team late and showed excellent content knowledge among all the team members. However, the students together formed some unique team characteristics. The Alpha members appeared to have strong interests in exploring the ST technologies and therefore practiced most of the tools at the beginning of the semester. The ST tools, especially the Whiteboard features, were used as collaboration tools and also used for fun activities. For instance, team Alpha members sometimes use color pens to draw cartoon images or play games (e.g., tic-tac-toe) during the off-task periods. Therefore, compared to the Gamma team, they were observed to be able to choose more effective tools or use multiple tools in their practice. For example, they used rectangle tool to build tables or used color tools to differentiate or highlight key points. At another side, the team appeared to easily lose focus and their meetings therefore resulted in low productivity. This may be explained by: (a) more technical issues encountered by the team increased the average meeting time which may bring up meeting fatigue; (b) side-talks, off-track conversations, or jokes that distracted the team attention and delayed work progress.

Gamma team examples: Personality differences among the Gamma team members may explain their different levels of participation. GL and BZ seem to be more talkative and social. They usually appeared to dominate most of the team conversation and maneuvered the direction of the meetings. BK seems to be very confident in the content knowledge and he was very conversant when he started talking about the content and his analysis work. It was difficult to make any judgment about MW because he rarely spoke during the meetings. Although every person was observed having distinctive personalities, the Gamma team appeared to have different *teamality* from the Alpha team. The team was very task-focused. Members were rarely observed in side-talks or joking. The team seemed satisfied with basic communication features (e.g., audio). Therefore, the team used fewer AIDE features, encountered much fewer technology issues, and appeared to have more efficient and effective meeting sessions and produced higher-quality performance.

Other Differences

Two other factors seemed to contribute to the team dynamics and may have influenced team communication pattern: content and institutional differences.

Content knowledge differences: The students who had better content knowledge appeared to engage more in team conversations. Although these students may show different personalities, they appeared to be more confident, conversant, and prepared, often raising good questions about particular points and, in most situations, dominated in the entire meeting. For instance, AF in the Alpha team appeared to be a good leader steering the team in proper directions. BK in the Gamma team always participated at the key times providing insightful thoughts, clarifying explanation and ideas. These students were observed less engaged in the off-task activities and chatter. This trend was observed throughout the entire semester.

Institutional differences: The institutional differences were distinctive between the Gamma team members. The two CU students focused more on problem solving, searching and utilizing more resources outside of the course to help with their analysis process. Meanwhile, the two SU students were more concerned about the timeline of completing the design project and appeared to have more team management skills. Skills from students from the two universities contributed to the success of the team performance.

Overall Evaluation of Team Communication

Overall, three major team communication themes emerged from the data analysis. First, different personalities, institutional cultures, content knowledge and skills contributed to the dynamics of each team. Second, data suggested from the cognitive imprinting theory perspective (Geer, 2005), a team's initial communication patterns were powerful predictors in determining subsequent interactive behaviors. However, the data also suggests that the team's communication pattern may change when the team dynamics change. For example, when the GL was not present, BK became conversant. He and BZ were observed dominating the conversation. Third, an explanation for AF's minor engagement in meetings still remains mystery at this point. Data collected from the sample archival video data do not suggest rationale for his limited engagement. However, when GL and BZ called MW several times asking for his thoughts about the design project MW rarely was able to provide a complete answer. Perhaps MW did not have sufficient content knowledge or he was not motivated to participate. There were also several occasions when MW was not in the meetings and GL criticized his low quality work commenting that this was 'the crappiest work he had ever seen.' Other members did not make specific comments, however, seemed to agree with GL. Thus, dynamics of the team, cognitive and initial media preferences, and unidentified individual characteristics were related to communication effectiveness and team productivity.

CONCLUSION

Team Communication vs. Team Technology Use and Adoption

This study helps extend the media stickiness theory by discovering that a person's choice of a particular technology within a virtual multiuser learning environment and sticking to its use relies primarily on needs related to communication, socialization, or problem solving. This perspective aligns with Geer and Barnes' (2007) suggestion that "…high levels of creativity through collaboration … are not always necessary to their immediate goals … the learning of participants needs to be considered." (p. 135). When a team's communication needs are satisfied, motivation for using other new tools is limited therefore tools deemed useless are eventually dropped suggesting the team does not see the necessity of using the tool regardless of how well it was designed to meet a communication need. For example, the Gamma team was observed sticking to the pen tool to draw tables rather than using the rectangle tool. This may be explained by the fact that the pen was sufficient and easy to use and use of the rectangle tools was more time consuming. Changing to a new way of working required a cost-benefit analysis (Huysman et al., 2003) suggesting that the cost of using the rectangle tools was more than using the pen tool in this case. Students' communication needs played a major role in the variation of tools they use and how long they continue to use each tool.

Adopting particular technologies and sticking to their use are not simply a team or any particular individual's decision. A team's selection of specific technologies tends to have the following characteristics: technologies frequently used in team members' daily life are more often adopted by the team and personal interests in technologies intrigue members into practicing additional new technologies but does not necessarily lead to the adoption of a particular new technology. The second characteristic may conflict with what Huysman et al. (2003) suggested that "lack of internal communication process was one potential source of media stickiness in virtual teams" (p. 432). JD in the Alpha team suggested using the typing tool instead of pen writing at the beginning of the semester. Some of the team members followed him using the typing tool, however typing was eventually dropped while pen writing remained a common activity. External feedback does not seem to promote students' adoption of particular tools either. The two instructors commented several times that neither team appeared to use the SameTime technology adequately although they had sufficient technology competences. However, the feedback criticizing these teams' habitual way of working did not have enough power to make a permanent change in both teams' technology use.

Two possible explanations for the media stickiness observed in this study are: 1) students may realize that use of more tools causes more technical issues and 2) using many tools reduced the teams' productivity and ability to meet assignment deadlines. Due to the busy schedule of the team members, the teams usually had one hour meeting times in which they needed to accomplish multiple tasks including technology normalization, scheduling, team logistics, and content discussion. Therefore experimenting with different new tools may have been deemed a low priority on the team's task list.

Team's technology use may relate to the team's communication characteristics. It was observed that the Alpha team, who practiced more technologies and encountered more technology issues, had lower productivity. While the Gamma team, using few technologies and rarely encountering technology issues, had much higher productivity and performed better in team management. This possible connection may suggest that the Alpha team did not have good team management skills to prioritize tasks. This also suggests that if the students face multiple learning goals (e.g., learning technology and learning content knowledge), they may not be able to do well in every goal if the task load is beyond their ability. As some of the students suggested that the course should be designed as one-year course because there were too many concepts to learn. Therefore, two instructional interventions to consider in the future course design include: 1) providing constant feedback in terms of team management including topics such as how to prioritize tasks and 2) balance instructional goals, student learning load, and the time allowed for completing the course.

Various factors contributed to team dynamics, which results in the formation of a team's 'personality' – or its *teamality*. A team's teamality may suggest a team's communication pattern, its technology choices, and the team management strategy. These patterns would remain consistent and / or grow stronger (e.g., MW in the Gamma team appeared to participate less and less in the meeting throughout the semester) as the theory of cognitive imprinting suggests. However, the data in this study also implies that when the team dynamics are altered due to some change in the team (e.g., more members joined the team, or someone is not present in the meeting), teamality would also change, which hence impacts the team's communication patterns and team productivity. For example, the Alpha team was observed becoming more productive after AF joined. In the future instructional design, it may be important to switch team members at the beginning of the course (e.g., in the Best Practices sessions) to help students learn about different personalities and communication patterns in virtual teams. Hence, they may be able to more quickly fit into a team collaboration culture and to develop more mature and insightful thoughts about how to learn with a team.

This study has revealed very specific details about how two distributed teams worked within a synchronous multiuser environment. The video analysis, surveying, and document analysis have allowed an intimate investigation of how and why learners use system and non-system technology based tools to collaborate, problem solve, and learn. These data suggested a 'teamality' variable influenced by the team dynamics, technology choices, and individual members' characteristics. Further unpacking of this concept may help in designing more powerful synchronous virtual learning environments and tools that help learners engage in deep learning and develop the competencies to be successful in multiuser virtual environments.

REFERENCES

Chen, G. (1994). Social desirability as a predictor of argumentativeness and communication apprehension. *Journal of Psychology Interdisciplinary & Applied*, *128*(4), 433–439. doi:10.1080/00223980.1994.9712749

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2007). Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education*, *49*(2), 309–329.. doi:10.1016/j.compedu.2005.07.003

Davidson, B. D., Davidson, R., Gay, G., Ingraffea, A., Miller, M., & Nozick, L. … Rath, C. (2002). Collaborative distance design of aerospace structures. *Proceedings of the 32nd ASEE/IEEE Frontiers in Education Conference,* Boston, Massachusetts.

Dieterle, E., & Clarke, J. (2005). Multi-user virtual environments for teaching and learning. In Pagani, M. (Ed.), *Encyclopedia of multimedia technology and networking* (2nd ed.). Hershey, PA: Idea Group, Inc.

Fåhræus, E. R. (1999). Tutoring group learning at a distance. *Proceedings of Society for Information Technology & Teacher Education International Conference*, *1999*, 152–157.

Geer, R. (2005). *Imprinting and its impact on online learning environments*. Australasian Society for Computers in Learning in Tertiary Education.

Geer, R., & Barnes, A. (2006). Media stickiness and cognitive imprinting: Inertia and creativity in cooperative work & learning with ICTs. *Proceedings of the World Computer Congress*, (pp. 55-64).

Geer, R., & Barnes, A. (2007). Beyond media stickiness and cognitive imprinting: Rethinking creativity in cooperative work & learning with ICTs. *Education and Information Technologies*, *12*, 123–136. doi:10.1007/s10639-007-9036-6

Gokhale, A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, *7*(1), 22–30.

Grabowski, B., & Koszalka, T. (2002). Helping educators harvest Internet resources: The development of technology integration support tools. *Proceedings of the IADIS International Conference WWW/Internet 2002*.

Hron, A., & Friedrich, H. (2003). A review of Web-based collaborative learning: Factors beyond technology. *Journal of Computer Assisted Learning*, *19*(1), 70–79. doi:10.1046/j.0266-4909.2002.00007.x

Huysman, M., Steinfield, C., Jang, C. Y., & David, K., Huis In 'T Veld, M., Poot, J., & Mulder, I. (2003). Virtual teams and the appropriation of communication technology: Exploring the concept of media stickiness. *Computer Supported Cooperative Work*, *12*(4), 411–436. doi:10.1023/A:1026145017609

Jonassen, D. H., & Kwon, H. (2001). Communication patterns in computer mediated versus face-to-face group problem solving. *Educational Technology Research and Development*, *49*(1), 35–51. doi:10.1007/BF02504505

Kapur, M., & Kinzer, C. K. (2007). Examining the effect of problem type in a synchronous computer-supported collaborative learning (CSCL) environment. *Educational Technology and Development*, *55*(5), 439–459. doi:10.1007/s11423-007-9045-6

Kirschner, P., Strijbos, J. W., Krejins, K., & Beers, P. J. (2004). Designing electronic collaborative learning environments. *Educational Technology Research and Development*, *52*(3), 47–66. doi:10.1007/BF02504675

Koszalka, T. A., & Wu, Y. (2008). *Evaluation of a cross-institutional collaborative distributed engineering educational environment*. 12th Global Chinese Computer and Communication Conference.

Lowyck, L., & Poysa, J. (2001). Design of collaborative learning environments. *Computers in Human Behavior*, *17*(5-6), 507–516. doi:10.1016/S0747-5632(01)00017-6

Nussbaum, E. M., Hartley, K., Sinatra, G., Reynolds, R., & Bendixen, L. D. (2004). Personality interactions and scaffolding in on-line discussions. *Journal of Educational Computing Research*, *30*(1-2), 113–136. doi:10.2190/H8P4-QJUF-JXME-6JD8

Ormrod, J. E. (2008). *Human learning* (5th ed.). Upper Saddle River, NJ: Pearson.

Panitz, T. (1996). *A definition of collaborative vs. cooperative learning*. Retrieved January 2, 2010, from http://www.city.londonmet.ac.uk/ deliberations/collab.learning /panitz2.html

Pellegrino, J. W. (2006). *Rethinking and redesigning curriculum, instruction and assessment: What contemporary research and theory suggests*. Retrieved February 24, 2010, from http://www.skillscommission.org/ pdf/commissioned_papers/Rethinking%20and%20 Redesigning.pdf

Prinsen, F., Volman, M. L. L., & Terwel, J. (2007). The influence of learner characteristics on degree and type of participation in a CSCL environment. *British Journal of Educational Technology*, *38*(6), 1037–1055. doi:10.1111/j.1467-8535.2006.00692.x

Rogers, E. M. (1983). *Diffusion of innovation* (3rd ed.). New York, NY: Macmillan.

Ross, J. A. (1996). *Computer communication skills and participation in a computer-mediated conferencing course*. Annual Conference of the American Educational Research Association.

Rourke, L., & Anderson, T. (2002). Exploring social communication in computer conferencing. *Journal of Interactive Learning Research*, *13*(3), 259–275.

Shuman, L. J., Atman, C. J., Eschenbach, E. A., Evans, D., Felder, R. M., & Imbrie, P. K. … Richard, L. G. (2002). The future of engineering education. *Proceedings of the 32nd Annual ASEE/IEEE Conference on Frontiers in Education*.

Solimeno, A., Mebane, M., Tomai, M., & Francescato, D. (2008). The influence of students and teachers characteristics on the efficacy of face-to-face and computer supported collaborative learning. *Computers & Education*, *51*(1), 109–128. doi:10.1016/j.compedu.2007.04.003

ENDNOTES

1 Team members took turns as the team moderator. The responsibilities of the meeting moderator included: (a) scheduling meetings and notifying the team (b) checking the ST technology to ensure it was functioning normally before the meeting; (c) uploading meeting agendas, assignments, and relevant documents to ST meeting before the meeting; (d). recording the meeting; (e). closing the recording and meeting

2 The Delta team was dismissed because three students dropped from the course. The composition of each team was reshuffled and AF was assigned to the Alpha team.

3 GL and MW were from SU and BZ and BK were from CU

This work was previously published in Multi-User Virtual Environments for the Classroom: Practical Approaches to Teaching in Virtual Worlds, edited by Giovanni Vincenti and James Braman, pp. 366-385, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 10
Modes of Open Innovation in Service Industries and Process Innovation:
A Comparative Analysis

Sean Kask
INGENIO (CSIC-UPV), Spain

ABSTRACT

This broad study empirically compares the returns to different open innovation approaches, namely forms of pecuniary acquisition and non-pecuniary sourcing, on both product and process innovation in low-tech service and manufacturing firms. A fixed-effects analysis reveals differing patterns of the effectiveness of open innovation strategies across sectors and type of innovation outcome, along with decreasing returns from being "too open". In general, the purchase of intangible intellectual property and broad search breadth have greater effects on product innovation, whereas the returns to knowledge embodied in physical artefacts and to drawing deeply from external sources are greater for process innovation. Overall, external sources of knowledge more strongly predict innovation in low-tech service firms than in the manufacturing sector. The final section considers implications for managers and policy makers.

INTRODUCTION

Open innovation, which posits that firms should use external knowledge in their internal innovation process, is an approach which is increasingly embraced by firms. In the short time since the term *open innovation* has been coined (Chesbrough, 2003b), numerous academic research projects, conferences, and specialized service providers quickly sprung up dedicated to the topic (Fredberg, Elmquist, & Ollila, 2008; Lichtenthaler, 2011). Many corporations have recently started to formalize open innovation into designated departments and roles; consider, for example, Hewlett-Packard's "Open Innovation Office" or

DOI: 10.4018/978-1-4666-1945-6.ch010

the employees with the job title "Director of Open Innovation" walking the halls of General Mills, Nokia, and Unilever. A number of leading firms have introduced open innovation competitions and initiatives, such as Sarah Lee's Open Innovation Portal or Cisco's I-Prize competition (Drakos, 2008). Indeed, authors have observed a bandwagon effect as open innovation gains momentum, with many senior executives "under increasing pressure to justify their refusal to cooperate with the outside world and exploit the open innovation wave" (Gassmann, Enkel, & Chesbrough, 2010: 215). Supported by numerous case studies that corroborate the positive results of open strategies (Chesbrough & Garman, 2009; Huston & Sakkab, 2006; Rohrbeck, Hölzle, & Gemünden, 2009), open innovation has become an imperative for firms over a relatively brief period of time.

However, despite all the excitement as policy makers and firms race to embrace open innovation, most of the research is drawn from case studies on product development in large, multinational high-tech firms or niche business models and open source (Chesbrough & Crowther, 2006). This means that, as a young, emerging theory, it lacks empirical support and generalizability across diverse conditions and businesses (Van de Vrande, Vanhaverbeke, & Gassmann, 2010; West, Vanhaverbeke, & Chesbrough, 2006: 302).

This chapter looks at some of these understudied circumstances by addressing different sectors, modes of open innovation, and objectives. In particular, it (1) compares the effectiveness of various inbound open innovation activities between low-tech manufacturing and service sectors; (2) evaluates the returns to several strategies for conducting open innovation, including different forms of pecuniary (monetary) and non-pecuniary external sourcing; (3) and examines objectives beyond products to include a firm's process innovation. The chapter also includes a discussion and analysis of the potentially adverse consequences of being "too open". The analysis links these factors to firm innovation performance by comparing the "returns" or "effect sizes" across innovation type and sectors. Thus the general aim is to bring these important yet understudied topics into the open innovation discussion. A panel survey of innovation activities in 3,800 Spanish low-tech manufacturing and service firms over a four year period provides the sample for the analysis.

The remainder of the chapter is organized as follows: (1) the first section reviews the literature and generates several hypotheses. This includes a review of process innovation, the nature of innovation in the service industries, the relationship of different modes of inbound open innovation with these factors, and decreasing returns from over-searching. (2) The empirical section presents the data and variables used in the analysis. It explains the methodology and advantages of fixed-effects estimation. (3) The next section discusses the results of the empirical estimations in relation to the hypotheses and the significance of the uncovered patterns. (4) The concluding section proposes implications for managers and policy makers and identifies directions for future research.

THEORY DEVELOPMENT AND HYPOTHESES

Product and Process Innovation

To date, the open innovation literature has focused on the generation of commercial products. Unfortunately, this concentration on product development neglects an important avenue of firm competitiveness and profitability: process innovation (Niehaves, 2010). Process innovation is defined as "new elements introduced into an organization's production or service operations—input materials, task specifications, work and information flow mechanisms, and equipment used to produce a product or render a service—with the aim of achieving lower costs and/or higher product quality" (Reichstein & Salter, 2006). Process innovation is an important source of firm performance and productivity growth (Vivero, 2002), and new processes used to produce and deliver goods and services are often central to a firm's ability to compete in terms of costs, quality, and flexibility (Pisano, 1997). The following section links differences in product and process innovation to particular open innovation activities.

Further to simply including process innovation, the structure of the analysis enables a distinction between process innovation in service and manufacturing sectors. Process innovation in service industries can differ substantially from that in manufacturing since, often by nature, the delivery of the service is tied to the locus of its production (Berry, Venkatesh, Parish, Cadwallader, & Dotzel, 2006). Competitive advantage comes not merely from *what* a service firm does for a customer, but *how* it is delivered. Since some authors argue that services are commoditized at an even faster rate than most goods (Pine & Gilmore, 1999), even low-tech service firms must continuously innovate in order to remain competitive. Therefore, returns to sourcing strategies on process innovation are considered and compared between industries in the analysis.

Modes of Inbound Open Innovation

Innovative search describes the adaptive process by which organizations learn, develop and explore new technologies and new ways of doing things (Levinthal & March, 1981). Search efforts can be inwardly (e.g. R&D, operations analysis) or externally (e.g. imitation, external sourcing) focused (Nelson & Winter, 1973), can vary in depth and scope of information sources accessed (Katila & Ahuja, 2002), and can be directed towards sources both within and external to a firm's industry (Katila, 2002). Each search strategy can thus vary along many dimensions such as locus, timing, knowledge source, and the degree to which limited resources are directed towards exploring new ideas or exploiting existing knowledge (March, 1991). Chesbrough's argument for the shift towards an open innovation paradigm predicts search strategies that are more externally oriented. As such, there is variety in potential external search strategies available to firms. More generally, firms differ in the level and nature of "openness", defined as "the degree to which the firm seeks to draw in new knowledge and to re-use existing knowledge from external sources" (Laursen & Salter, 2004). Accordingly, different search strategies result in differing degrees of innovative performance depending on the nature of the innovation in question.

Although much of the original rationale for open innovation revolves around inadvertent spill-overs and missed opportunities to commercialize technologies developed in-house, or "inside-out" open innovation (Chesbrough, 2003b; Chesbrough & Rosenbloom, 2002), the research and practitioner interest to date has focused on the benefits of "outside-in" open innovation. This "inbound" open innovation involves identifying and appropriating external knowledge for commercialization or implementation. Two meaningful dimensions of inbound open innovation are sourcing (non-pecuniary) and acquiring (pecuniary) (Dahlander & Gann, 2010).

Acquiring involves the procurement and purchase of intellectual property (IP) or artefacts developed outside the firm. Firms may buy external technology for any number of reasons, such as insufficient internal R&D resources, diversification into new competencies, or simply to access cutting-edge, specialized technologies developed by other organizations (Lowe & Taylor, 1998). Since some pecuniary, commercial transaction takes place, acquiring infers that the firm foresees some immediate application for the externally-sourced knowledge artefact. In other words, since it is unlikely that a firm would expend financial resources simply for the purposes of exploration, acquiring occurs with exploitation and direct application or commercialization in mind. I further consider two forms of pecuniary open innovation: the acquisition of *intangible* (or disembodied) knowledge in the form of patents, licences, or other IP, and the purchase of tangible knowledge *embodied* in physical artefacts such as machinery, equipment, and software (Baetjer, 2000; Vega-Jurado, Gutierrez-Gracia, & Fernandez-de-Lucio, 2009).

The open innovation literature advises firms to seek out opportunities to licence in technologies, namely in the form of patents and IP, which can be commercialized as products in the firm's current business model (Chesbrough, 2006). Previous research, for example from the Finnish manufacturing sector, indicates that use of intangible knowledge is more often applied to product innovations whereas process innovation results from the incorporation of knowledge embodied in artefacts (Rouvinen, 2002). When a firm incorporates a new piece of machinery or software (embodied knowledge) into the production process, one firm's product becomes an input into another firm's processes. The incorporation of knowledge embodied in equipment into internal infrastructure allows firms to "know less than they buy" while still achieving improvement in processes (Flowers, 2007). Regarding intangible technological IP, since the internal processes of most firms remain secretive and proprietary, there is less of an incentive or market for firms to purchase intangible IP inputs into processes. Thus there is a distinction between purchasing components for a production process (buy to build) and purchasing capital goods for operational infrastructure and processes (buy to use). I draw the following hypotheses from this distinction between intangible and embodied knowledge:

H1a: *The effect of pecuniary acquisition of intangible technological IP is greater for product innovation than for process innovation.*

H1b: *The effect of pecuniary acquisition of embodied knowledge is greater for process innovation than for product innovation.*

In contrast to acquiring, *sourcing* entails scanning the environment for information related to new technologies and new opportunities, which the firm may then decide to act on, whether through instigating internal R&D or acquiring external technologies (Chesbrough, Vanhaverbeke, & West, 2006). The constructs *breadth* and *depth* (Laursen and Salter, 2006) are roughly analogous to external search scope and depth (Katila & Ahuja, 2002) but more precisely describe the variety and intensity of external sources from which organizations draw knowledge. These constructs thus represent variations in external sourcing strategy. An interesting aside from Laursen and Salter (2006) is that depth is a stronger predictor of sales from radical product innovations, whereas the effect of breadth is greater for sales from incremental product innovations. That is, the form of sourcing strategy impacts product innovation differently.

Some authors differentiate two forms of process innovation: *technological process innovation,* which results from the incorporation of new goods and embodied knowledge into production or service delivery, and *organizational process innovation*, which involves learning and more intangible changes in the coordination of human resources, such as new management practices or ways of doing things (Edquist, Hommen, & McKelvey, 2001: 14-17). The prediction in Hypothesis 1 (that the incorporation of embodied knowledge has a greater effect on process innovation than product innovation) follows from technological rather than organizational process innovation. Learning from external, non-pecuniary sourcing, such as through best-practice sharing at conferences, management training from universities, or working closely with clients to better serve their needs, can also influence organizational process innovation; external sourcing, and not just pecuniary inbound innovation, should influence process innovation, and different sourcing strategies, namely breadth and depth, can affect process innovation in different ways. For starters, the organizational changes required for process innovation are not easy to come by, with an emerging literature on the difficulties of organizational process change management (Kettinger & Grover, 1995). Organizational processes are intangible and often tacit in nature, requiring a fair amount of effort to understand and explicate, develop or imitate, and implement. Therefore, deeper, more intense relationships with external sources are needed to adequately develop, transfer, or implement organizational process innovations. Conversely, breadth of search provides firms with a wider yet shallower scope of available knowledge and opportunities which the firms may decide to further develop into products (Laursen & Salter, 2006). These differences in the nature of sourcing related to product and process development lead to the following hypotheses:

H2a: *The effect of non-pecuniary search breadth is greater on product innovation than on process innovation.*
H2b: *The effect of non-pecuniary search depth is greater on process innovation than on product innovation.*

Open Innovation in Service Industries

Despite making up significant portions of the economy, there is a relative paucity of research on innovation in service industries, not to mention in an emerging theme such as open innovation. Many leading firms currently have very different ways of defining service innovation and the processes which enable it (Anderson-Macdonald & Kask, 2010). However, despite the "manufacturing bias" in innovation studies (Drejer, 2004), there is good reason for the growing interest in the antecedents and outcomes of service innovation: developed countries are predominantly moving towards service economies, with the sector comprising well over 70% of the share of economic activity (Anderson, Howells, Hull, Miles, & Roberts, 2000; Boden & Miles, 2000; Gallouj & Djellal, 2010; OECD, 2000; Rubalcaba, 2007). The analysis of the consequences of growth in services – and corresponding reduction in the proportion of the labour force engaged in manufacturing – has a long history of theoretical and political debate, with views of the service sector ranging from a passive adopter of technology and "innovation laggard" (Baumol, 1967) to "the core engine of the new knowledge-based economy" (Gallouj & Savona, 2009). Whatever the impact of services on the economy as a whole, services represent a significant area of activity which has largely been left out of the open innovation discussion.

How might innovation in services differ from the innovation traditionally studied in manufacturing? Some authors propose a radically different model of innovation in services, suggesting a "reverse product cycle" whereby incremental improvements first increase the efficiency of delivery of existing services, only later to lead to improved service quality through process innovations and finally to the generation of new service products (Barras, 1986, 1990; Gallouj, 1998). Non-technological and tacit knowledge play an important role in service innovation (Tether et al., 2007), and informational activities within service firms, largely facilitated by ICTs and the network of actors and interactions between them, lead to knowledge creation that is not bound within specific domains in traditional R&D labs (De Bandt and Dibiaggio, 2002). This suggests that, although service industries share as many similarities as differences with manufacturing when it comes to innovation, interaction with the external environment is crucial for service innovation.

In light of this, the model compares the returns to open innovation activities between service and manufacturing sectors. Although some research has shown that service and manufacturing firms often exhibit comparable patterns of the use of knowledge sources for technological innovation (Sirilli & Evangelista, 1998), the relative returns to these sources may differ substantially between industries. If service firms benefit from interactions with external suppliers due to the nature of service innovation and are indeed more passive adopters of technological innovation, then inbound open innovation will more strongly predict innovative performance in service firms than in the manufacturing sector. This leads to the following hypothesis:

H3: *Overall, the effect of inbound open innovation activities are greater in low-tech service firms than in low-tech manufacturing firms.*

Decreasing Returns to Openness

Finally, although the hypotheses in this chapter and evidence from previous studies predict a positive effect of external knowledge on both process and product innovation, one highly-cited contribution proposes an inverted-u shaped relationship between the number of kinds of external sources and innovative performance of products (Laursen & Salter, 2006). Using a cross-sectional survey of British manufacturing firms, these authors find that firms drawing from external knowledge sources will have a better innovative performance than firms relying solely on themselves – up to a point. After a certain level, firms experience decreasing returns to external sourcing. Graphically, the relationship between innovative performance, in this case measured as the percentage of sales from new products, and breadth or depth takes on an inverted u-shape. A more recent, similar study using a slightly different econometric methodology looks at Finnish manufacturing firms, again finding strong support for the benefits of external search breadth but with some evidence for decreasing returns (Leiponen & Helfat, 2010). How is it that firms can be "too open"? Laursen and Salter posit that decreasing returns set in because of (1) natural constraints on absorptive capacity, (2) Not Invented Here (NIH) syndrome, (3) timing, and (4) limited allocation of managerial attention.

First, absorptive capacity is the ability to recognize, internalize, and exploit external knowledge to commercial ends (Cohen & Levinthal, 1990). Building absorptive capacity is seen as a prerequisite for firms to successfully engage in inbound open innovation (Spithoven, Clarysse, & Knockaert, 2010). Conceptualized as a *capacity*, it has limits. When a wide array of diverse external knowledge is accessed, much of that knowledge may or may not be appropriate or relevant to the firm. Sorting out which knowledge is relevant requires resources and great effort. Simply put, there may be too many ideas to comprehend and utilize at one point in time, leading to "information overload." Similarly to the absorptive capacity problem, attention-based theory of the firm places limits on the attention of managers (Simon, 1997). According to this theory, managerial attention is the most valuable resource in a firm, and the behaviour of the firm results from how decision makers focus and distribute their limited attention (Ocasio, 1997). The allocation of attention, especially from senior management, during the innovation search process is important for performance (Koput, 1997; Yadav, Prabhu, & Chandy, 2007). Managers who are bombarded by information have little time to contemplate it all and may direct their attention to "safe" projects.

"Not Invented Here" (NIH) syndrome, which the open innovation literature mentions often, is the rejection of knowledge and technologies originating externally to the firm, which can be detrimental to firm performance (Katz & Allen, 1982). Some level of scepticism stemming from the uncertainty of externally acquired knowledge can be healthy, but too much can be a major barrier to open innovation (Chesbrough, 2003b: 182). As the amount of external knowledge being accessed increases, the effects of NIH syndrome can become more apparent. Although not typically addressed in open innovation studies, the case also exists where firms overvalue external knowledge, which may lead to detrimental effects and poorer innovative performance (Lichtenthaler & Ernst, 2006). The "Not Invented There" (NIT) syndrome is defined as occurring when "people in management show more interest in what is going on elsewhere than in their own laboratories" (Laden, 1996). For NIT syndrome, "the other person's dessert always looks better", which causes management to spend more time evaluating external sources because they may already be technically feasible or on the market. Menon and Pfeffer demonstrate empirically the prevalence of some firms having a preference for external knowledge over their own internal knowledge (2003). This is due to (1) status associated with gaining knowledge from external sources and (2) the fact that external knowledge is rarer which makes it appear special and unique. Firms at a higher degree of external search breadth may be over-zealous and overly positive about drawing from external knowledge sources, resulting in inadequate focus on internal capabilities and mistaken selection of inappropriate knowledge. Thus the poorer innovation performance relative to the extremes of external sourcing may reflect both sides of the NIH-NIT coin.

Regarding the "timing" problem, too many ideas may come at the wrong time to be fully exploited. With a lot of potentially good ideas coming at once, the ratio of good ideas passed over is higher. Chesbrough touches upon a similar reason regarding timing, where accelerated cycle times for projects allows firms to assess only so many ideas or technologies at a time, leading them to eventually pass over some (Chesbrough, et al., 2006: 17). Some innovations take years before a use is discovered for them. For example, one of the inventors of the laser, Gordon Gould, said that for the first five years after its discovery, the laser was "a solution in search of a problem" (Brown, 1988: 310). When companies are inundated with ideas, some may get buried when the appropriate time finally emerges.

Research by Chandy et al. similarly finds this inverted u-shaped relationship between a firm's ability to convert ideas into innovations: "a strong focus on speed and on generating many ideas may actually hurt firms by lowering their conversion ability" (2006). They hypothesize that problem solving ability is influenced by four factors, namely (1) workload, (2) time pressure, (3) expertise, and (4) task importance. At high levels of external search breadth and depth, firms may be hurting their conversion ability, resulting in decreased performance. This paradox of allotting the right level of pressure in the open innovation process has also been noted in practitioner case studies (Mesaglio & Hunter, 2008).

Therefore, consistent with Laursen and Salter (2006), the following hypothesis predicts an inverted-u relationship between external sourcing and the performance of both product and process innovation:

H4: *non-pecuniary breadth and depth are positive predictors of both product and process innovation up until a point, after which firms experience decreasing returns to external sourcing.*

DATA AND EMPIRICAL ESTIMATION

Data Source

The analysis in this chapter is carried out using the PITEC (*Panel de la Innovación Tecnológica*), which is based the OECD's Oslo Manual. The survey is administered by a joint effort of the Spanish National Statistics Institute (INE), the Spanish Foundation for Science and Technology (FECYT), and the Foundation for Technical Innovation (COTEC). The unit of analysis is the single enterprise, whether part of a group or independent. The PITEC differs from most other OECD-based innovation surveys in two important and favourable aspects. First, firm participation is mandated by law (*Leyes* 4/1990, 13/1996, and article 10.1 of the LFEP), which limits problems associated with respondent selection bias; since many of the national innovation surveys in other countries are voluntary, this opens the door to respondent selection bias with only eager, innovative firms responding. Second, as the "P" in the acronym indicates, it is structured as a panel dataset, with observations repeated on the same firms over time. As discussed below, there are several advantages to panel data.

The dataset contains observations on more than 2,200 low-tech manufacturing and 1,600 low-tech service firms over four periods, with the most recent ending in 2007. Firms are assigned to 31 sectors according to the Spanish National Classification of Economic Activities (CNAE-93 - *Clasificación Nacional de Actividades Económicas*), roughly equivalent to the NACE system. The sectors are divided into high-tech/low-tech manufacturing and services along this classification system by the INE. The specific industries are included in Table 1.

Table 1. Manufacturing and service industries Included in analysis

Low-Tech Manufacturing	Low-Tech Services
Food and Beverages	Electricity, Gas and Water Utilities
Tobacco	Construction
Textiles, Tailoring and Furs	Sales and Repair of Automobiles
Leather and Footwear	Wholesale
Lumber and Cork, Paper	Retail
Editing, Arts, Graphics, and Copying	Hotel
Petroleum Refinement	Transport
Rubber and Plastic Materials	Transport, Travel Agencies
Glazed and Tile Ceramics	Financial Intermediation
Non-Metallic Mineral Products	Real Estate
Ferrous Metallic Products	Architectural services
Non-Ferrous Metallic Products	Testing and Analysis
Metallic Products (excluding Machinery)	Other Business Services
Shipbuilding	Education
Furniture	Film and Video Activities
Toys and Games	Radio and Television
Recycling	Other Sanitary, Social and Collective Services

Exploiting Panel Data: Accounting for Unobserved Firm Heterogeneity

Undoubtedly, there are many idiosyncratic and contextual factors which influence a firm's behaviour and ability to innovate, not all of which can be measured and included as variables in the analysis. Some obvious examples are star managers, proximity to resources such as clusters, organizational structure, etc. These are collectively termed "unobserved heterogeneity" in the management literature, and failing to account for it can result in bias due to omitted variables. Fortunately, the properties of panel data – repeated observations on the same firm over time – allow one to account for much of this heterogeneity via *fixed-effects models*. Fixed effects are factors at the level of the individual firm which remain more-or-less constant over time. In essence, each individual firm acts as its own counterfactual before and after some "treatment" of interest, in this case the use of various sources of knowledge. Because fixed effects models estimate variation in a firm over time, it is also called the *within* estimator. The inclusion of time and controlling for unobserved, omitted variables gets us closer to determining causal effects (Angrist & Pischke, 2009: 115-117). Although fixed-effect models are less efficient (i.e. it is difficult to achieve statistical significance), the estimates are consistent (i.e. the estimated coefficients are close to the "true" value). For these reasons, I employ a fixed-effects model described in the following description of the estimation procedure.

Dependent Variables: Measures for Effect of Product and Process Innovation

The dependent variables evaluate the effect of innovation on each firm's products and processes. The survey measures the *effect of product innovation* along three dimensions: (1) improved range of products or services; (2) penetration into new markets or increased market share; and (3) improved quality of product or services. The dependent variable, PROD_EFF, counts the number of effects rated as "high impact" and, therefore, ranges from 0-3. Firms rate the *effect of process innovation* along three dimensions which are relevant to both product and service firms: (1) improved flexibility of production or service provision; (2) improved capacity of production or services; (3) reduced labour costs. The variable PROC_EFF ranges from 0-3 where firms rated the impact of innovation on processes as "high".

Covariates

Two dummy variables measure pecuniary inbound open innovation: TECIP and EMBK. TECIP indicates whether the firm purchased external technological IP, such as patents or licences, during the period. EMBK indicates the purchase of external knowledge embodied in physical artefacts, such as in machinery, equipment, or software, used for product or process innovation. Finally, the log of R&D intensity, LRDIN, accounts for the internal resources used for innovation, measured as the spending on internal R&D as a percentage of total firm revenue.

NPBREADTH measures the number of kinds of external knowledge sources used for innovation which are considered *non-pecuniary*. This is similar to the *breadth* construct used by Laursen and Salter (2006), but excludes *suppliers* and *knowledge-intensive business services* as categories because of the potential to confound free or low-cost external public knowledge sources with pecuniary open innovation – firms handsomely pay suppliers and consultants, whereas fees for things like conferences or access to publications, if any, are negligible. The eight remaining types of external source include: clients, competitors, universities, public research organizations, technological centers, conferences and fairs, journals/publications, and professional associations. Likewise, NPDEPTH is the number of types of non-pecuniary external sources from which the firm has drawn intensely for innovation. Note that in order to test for decreasing returns and the inverted-u shape relationship predicted in Hypothesis 4, the regression models include the squared terms NPBREADTH_2 and NPDEPTH_2.

The within-regression accounts for variables that are subject to change in an individual firm over time, with constant variables accounted for with the fixed effects. This parsimonious approach thus excludes many unchanging characteristics that differ between firms and are normally included as controls in cross-sectional analyses. I do however include three variables as controls for changes in firm strategy or circumstances. First, LOGEMP, the log number of employees, controls for changes in firm size due to things such as growth, acquisitions, mergers, divestitures, etc. Second, since firms operating in international markets often face more diversified competition and have a higher propensity to innovate (Frenz, Girardone, & Ietto-Gillies, 2005), GEOMARKET indicates the geographical market in which the firm sells its products or services, taking on the values 1 (local market only), 2 (national, within Spain), 3 (inside the European Union), or 4 (international, outside the EU). Finally, dummies for each year in the panel control for cross-industry, macroeconomic fixed effects. Table 2 lists the descriptive statistics by industry for the variables before log transformation. With the exception of surprisingly higher R&D intensity and larger size of service firms, the dependent and key sourcing variables show no significant differences between industries.

Estimation Procedure

The dependent variables are discrete integers in panel data format, making a conditional fixed-effects (CFE) Poisson model appropriate (Hausman, Hall, & Griliches, 1984). The Poisson is attractive because it allows for a large number of zero values in the dependent variable and per-

Table 2. Descriptive statistics by industry for innovation-active firms, prior to log transformation

Variable	Range	Average (standard deviation)	
		Manufacturing	Services
Product Innovation	0-3	1.06 (0.616)	0.858 (0.545)
Process Innovation	0-3	0.575(0.575)	0.491 (0.512)
Non-Pec. Breadth	0-8	4.37 (1.62)	4.09 (1.57)
Non-Pec. Depth	0-8	0.765 (0.683)	0.735 (0.672)
Internal R&D	0-100	1.75 (2.98)	3.71 (4.28)
Technological IP	0/1	0.052 (0.165)	0.085 (0.199)
Embodied Knowledge	0/1	0.337 (0.375)	0.334 (0.356)
Size: # employees	10-41,509	169 (79.6)	637 (473)
Geographical Market	1-4	3.36 (0.348)	2.43 (0.363)
Year dummies	-	-	-

forms well in the presence of unknown forms of heteroskedasticity. Although normally used for discrete data counting the occurrence of some event, the estimator is consistent as long as the conditional mean is correctly specified, meaning that the dependent variable doesn't even need to be an integer (Santos Silva & Tenreyro, 2006). As per the usual convention, the variables are log transformed in order to provide a better fit with the distribution of the Poisson, which stipulates that the mean be equal to the variance lest the distribution is over-dispersed (Gourieroux, Montfort, & Trognon, 1984). Furthermore, these log-transformed variables are dimensionless and in the same range, enabling the comparison of effect sizes between product and process innovation. The CFE Poisson model drops those observations with either all-zero outcomes or unchanging sets of independent variables over time, which could introduce a systematic selection bias (Reitzig & Wagner, 2010). However, in the models run in this study, there is sufficient variability in the independent variables such that all innovative firms are kept in the analysis. In any case, fixed-effects models are robust against most kinds of selection bias since the selection characteristics are largely controlled for in the fixed effects (Kennedy, 2003: 312), such that the analysis estimates the variables within the firms of interest. Stata 11 executes the CFE Poisson regressions with the robust standard errors, which produces standard errors in line with more complex repeated-sampling procedures (Cameron & Trivedi, 2010).

The magnitude and significance levels of the coefficients tell an interesting story if one makes the comparison between industry, source of knowledge, and product or process innovation. Since the fixed effects controls for many idiosyncratic factors between low-tech manufacturing and service firms, we can compare the different estimated coefficients. A Chow test is normally the appropriate method in linear regression for determining if the coefficients of subsegments within a population, in this case low-tech manufacturing and service firms, differ statistically from one another due to some underlying structural break (Chow, 1960). However, a Chow test is not appropriate in the case of a non-linear maximum likelihood model such as the Poisson; instead, the better option is to run separate models on each group and compare the resulting estimated coefficients (Hoetker, 2007). Because of these "nonstandard standard error issues", I employ a non-parametric bootstrapping procedure on the difference in the estimated coefficients by accounting for the resulting residual variation matrix (Angrist & Pischke, 2009: 155-

164). Bootstrapping means simply repeatedly running the model on randomly drawn samples from the dataset (with replacement) so that the resulting standard error is the standard deviation of an estimator from these many draws. Although computationally burdensome, the advantage is that one does not make any assumptions about the underlying error distribution. The resulting p-values of the differences in coefficients tells us the magnitude and whether the coefficients for each variable are significantly different for the two industrial sectors. Using this procedure, Table 4 and Table 5 in the next section present the results of the comparison of coefficients across groups and type of innovation.

DISCUSSION OF RESULTS

Table 3 lists the results of the four CFE Poisson regressions. The same set of covariates is regressed on the variables for the effect of product innovation and process innovation in separate regressions for each of the service and manufacturing subsamples. With the exception of technological IP and the control variables, the covariates are all significant and positive in magnitude. The interpretation of individual coefficients in a Poisson model is as follows: the expected change in the (log) count of the effect of product or process innovation for a one-unit increase in the variable is equal to the magnitude of the individual coefficient.

Pecuniary Acquisition and Type of Innovation

Although a quick look at Table 3 reveals some interesting patterns in the returns to sourcing strategies across sectors and type of innovation, evaluation of Hypotheses 1-3 requires a statistical test to determine whether these differences are significant. To this end, Table 4 reports the results

Table 3. Conditional fixed-effects poisson regression results, by industry and type of innovation

	Product Innovation		Process Innovation	
	Manufacturing	Services	Manufacturing	Services
	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)
Non-Pec. Breadth	0.355*** (0.024)	0.463*** (0.032)	0.219*** (0.036)	0.331*** (0.041)
(Non-Pec. Breadth)^2	-0.032*** (0.002)	-0.042*** (0.003)	-0.016*** (0.004)	-0.030*** (0.009)
Non-Pec. Depth	0.157*** (0.025)	0.251*** (0.035)	0.240*** (0.041)	0.283*** (0.050)
(Non-Pec. Depth)^2	-0.017*** (0.005)	-0.027*** (0.006)	-0.026*** (0.008)	-0.029*** (0.009)
Internal R&D	0.081*** (0.019)	0.047* (0.025)	0.101*** (0.032)	0.079* (0.042)
Technological IP	0.010 (NS) (0.041)	0.125** (0.045)	0.043 (NS) (0.072)	0.030 (NS) (0.076)
Embodied Knowledge	0.078*** (0.017)	0.096*** (0.025)	0.191*** (0.032)	0.274*** (0.044)
Firm Size	0.121** (0.062)	0.016 (0.069)	-0.080 (0.091)	-0.011 (NS) (0.089)
Geographical Market	0.105*** (0.022)	0.013 (0.026)	0.075** (0.035)	-0.017 (NS) (0.038)
year dummies	included	included	included	included
# observations	8266	5790	5543	4259
# firms	2274	1625	1509	1180
Wald chi^2	493.68	458.74	257.26	281.37
prob> chi^2	0.0000	0.0000	0.0000	0.0000
LL	-4005.55	-2543.29	-2510.76	-1739.48

Significance levels: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$ (NS) Not Significant
Robust standard errors used.

Table 4. Statistical significance of difference between estimated coefficients

	Coeff. Product Innovation – Coeff. Process Innovation			
	Manufacturing Sector		Service Sector	
	Δ coeff. (s.e.)	p-value	Δ coeff. (s.e.)	p-value
Non-Pec. Breadth	0.137*** (0.041)	0.001	0.131*** (0.045)	0.003
Non-Pec. Depth	-0.084** (0.043)	0.038	-0.032 (NS) (0.058)	0.580
Internal R&D	-0.020 (NS) (0.036)	0.566	-0.032 (NS) (0.048)	0.508
Technological IP	-0.033 (NS) (0.080)	0.678	0.094 (NS) (0.079)	0.235
Embodied Knowledge	-0.113*** (0.038)	0.002	-0.178*** (0.051)	0.000

Significance levels: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$ (NS) Not Significant, Wald Chi^2 test
Bootstrap method with 400 repetitions.

of the bootstrapped standard error procedure described above in Estimation Procedure. The estimated coefficients for each variable's regression on product innovation are subtracted from the estimates for process innovation, by industry, in order to indicate the magnitude of the difference. The p-value tells us whether this difference in estimated coefficients is statistically significant.

Table 3 provides some support for Hypothesis 1a, whereby pecuniary acquisition of technological IP has a positive and significant effect on product innovation, however only for the low-tech service industry. Since technological IP is not significant in the case of process innovation or for product innovation in low-tech manufacturing firms, any comparison between industries is moot. The results in Table 4 strongly support Hypothesis 1b, indicating that returns to pecuniary acquisition of embodied knowledge are greater for process innovation than product innovation in both the services and manufacturing industries.

The finding that acquiring technological IP has no impact on process innovation nor on product innovation in low-tech manufacturing firms has intriguing implications for open innovation, especially considering the importance the open innovation literature places on inbound IP (Chesbrough, 2003a). These results echo findings from Tsai and Wang's study of the low-tech manufacturing sector, whereby technological IP sourcing had no predictive effect on more radical product innovation (2009). Similar results have been shown at the project level (Kessler, Bierly, & Gopalakrishnan, 2000). Other findings, however, indicate that sourcing technological IP becomes effective in high-tech industries as a firm's internal R&D capacities increase (K.-H. Tsai & Wang, 2007), suggesting that the impact of acquiring IP is a function of the technological intensity of the sector. Interestingly, in contrast to findings from the manufacturing sector, external IP sourcing seems to have a significant relationship with product development in low-tech *service* firms.

Non-Pecuniary Sourcing and Type of Innovation

Relatively strong support for Hypotheses 2a and 2b is evident in Table 4: the effect of non-pecuniary search breadth (2a) is greater for product innovation in both the service and manufacturing industries. On the other hand, returns to non-pecuniary search depth (2b) are significantly larger for process innovation, but only for manufacturing firms. This indicates that, at least in low-tech sectors, firms casting a wide yet shallow net (non-pecuniary breadth) are more likely to discover avenues for product innovation. This more superficial level of external sourcing may not be as appropriate for finding and implementing innovative processes. The comparable magnitude of the effect of non-pecuniary search depth on product and process innovation in service industries is intriguing. Although the effect of depth is slightly higher for process innovation in service firms, this difference is not enough to be statistically significant. This finding serves as evidence for dissimilar kinds of innovation processes between these two broad sectors.

Open Innovation in Manufacturing and Service Industries

In order to test Hypothesis 3, which predicts higher overall relative effect of inbound open innovation activities in service firms, Table 5 lists the results of the bootstrapping procedure on the difference in the estimated coefficients between industries.

Taken as a whole, the significant and positive results in Table 5 largely support the hypothesis of greater returns to inbound open innovation for service firms as compared to the manufacturing sector. For product innovation, the effects for the purchase of technological IP and non-pecuniary sourcing are significantly higher for service firms. The same trend is evident for process innovation in service firms. From Table 3 and Table 5 there is weak evidence that internal R&D is more important for manufacturing firms since the difference is not statistically significant. This is likely partly due to the markedly lower level of significance of the internal R&D coefficient for service firms, which in itself further supports evidence for lower returns to internal R&D in low-tech service firms. Despite the higher average R&D intensity in service firms evident in Table 2, the returns from internal R&D are lower than for manufacturing firms. The service firms seem to live up to their reputation as adopters of externally developed technology, although ultimately services get higher innovation returns to external sourcing than their low-tech manufacturing cousins.

Decreasing Returns from Being "Too Open"

Finally, Hypothesis 4 predicts decreasing returns to breadth and depth of non-pecuniary external sourcing. The significant and negative coefficients on the squared terms in Table 3 strongly support this. That is, an inverted-u shaped relationship exists between both product and process innovative performance and the breadth and depth of external sourcing. The relationship holds equally for service and manufacturing industries.

It is possible to roughly calculate the inflection point at which decreasing returns occur by setting the first derivative of the estimated coefficients equal to zero and solving. Given the scale for breadth and depth, which ranges from 0-8, on average decreasing returns set in for non-pecuniary breadth after about 5.5 external sources and for non-pecuniary depth at 4.5. Stated in a

Table 5. Statistical significance of difference between estimated coefficients

	Coeff. Service Sector – Coeff. Manufacturing Sector			
	Product Innovation		Process Innovation	
	Δ coeff. (s.e.)	p-value	Δ coeff. (s.e.)	p-value
Non-Pec. Breadth	0.107*** (0.039)	0.006	0.112** (0.057)	0.048
Non-Pec. Depth	0.094** (0.045)	0.027	0.043 (NS) (0.076)	0.542
Internal R&D	-0.034 (NS) (0.036)	0.339	-0.228 (NS) (0.058)	0.696
Technological IP	0.115* (0.062)	0.065	-0.012 (NS) (0.120)	0.917
Embodied Knowledge	0.018 (NS) (0.034)	0.588	0.083 (NS) (0.062)	0.183

Significance levels: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$ (NS) Not Significant, Wald Chi^2 test
Bootstrap method with 400 repetitions.

more meaningful way that is comparable and consistent with results from Laursen and Salter (2006), firms sourcing from more than two-thirds of the available external knowledge sources and using more than half of those sources intensely experience decreasing or even negative returns. Unsurprisingly, these results also indicate that decreasing returns set in much sooner for depth than breadth: depth requires higher involvement and quickly strains internal resources.

The recent enthusiasm surrounding open innovation and the detrimental effects of over-searching indicated in this analysis suggest that firms' external search strategies can be "too open". The presence of decreasing returns to external sourcing highlights the importance of neither being too open nor too closed to the external innovation ecosystem. That is, focused and targeted open innovation efforts lead to optimal returns.

FUTURE RESEARCH DIRECTIONS

Thanks to much attention from the academic research community, a fair amount of progress has been made in our understanding of open innovation over a relatively short period of time. However, as an emerging theory there are still many questions to explore. First and foremost, an objective of this chapter is to bring low-tech service industries and process innovation into the open innovation discussion. As such, this comparative analysis provides some initial clues about how open innovation differs across these subjects, but much further research is needed.

Innovation survey data is only able to capture the breadth and depth of different *categories* of external sources used for innovation. Although this is a good proxy for a firm's external sourcing strategy and level of openness, access to more granular data on the actual nature and number of actors and interactions with external sources of knowledge would provide a much richer understanding of the optimal external orientation of the firm. This kind of data could well be supplemented with a better understanding of the relationship between organizational forms and internal practices which influence a firm's ability to benefit from various open innovation strategies.

Although the analysis in this chapter focuses on search strategy regarding external sources of innovation, it does not address the myriad tools currently being proposed and used by firms generally included under the umbrella of open innovation. These include crowdsourcing competitions, the use of open source strategies, lead user methodologies, co-creation and mass customization. Although there is a growing volume of research into these specialized tools, many questions remain regarding their effectiveness and appropriateness across diverse contexts, such as various types of intended innovation outcomes, innovation of commercial services versus goods, or how these tools and techniques can be combined.

CONCLUSION AND IMPLICATIONS

This chapter addresses the following questions: How effective are various open innovation strategies (pecuniary and non-pecuniary) across sectors (manufacturing and services) and objectives (product and process innovation)? And how open should a firm be?

First, the literature and empirical analysis suggest that forms of pecuniary and non-pecuniary external sourcing are differentially effective according to innovation objective, namely product and process innovation. Predictably, external knowledge embodied in artefacts, such as equipment, components or even software, is effectively integrated into firms' processes but is a weaker predictor of product innovation. Although the open innovation literature places an emphasis on the acquisition of intellectual property, in this sample from low-tech industries, the acquisition of patents and other intangible IP did not predict process innovation and predicted product innovation only in service firms. This could be significant for low-tech firms in general; physical artefacts (and software systems) are "pre-packaged" knowledge that can be more easily exploited, whereas low-technology firms may face greater knowledge barriers to valuing and exploiting intangible intellectual property. Different approaches to drawing from non-pecuniary external sources of knowledge also impact a firm's propensity for product and process innovation. In general, greater "depth" of interaction is needed for process innovation, whereas "breadth" – drawing from a broad range of external actors – seems to be more pertinent for the effect of product innovations.

This chapter contributes empirical evidence for the effectiveness of knowledge sourcing in low-tech sectors. Similarly to the lack of research on service innovation, low-tech industries face a paradox, with these "traditional" industries making up the bulk of economic activity in many countries but still being largely left out of innovation studies (Hirsch-Kreinsen, 2008). Although any comparative analysis between high- and low-tech firms would be out of the scope of this chapter and no specific hypotheses are given regarding low-tech industries, it aims to serve as a starting point to bring these important sectors into the open innovation discussion. In general, the analysis reveals that external sources of knowledge have a greater impact on service-sector firm innovation in comparison with their low-tech manufacturing counterparts, who seem to benefit more from investments in internal R&D.

The arguments and evidence brought forth in this chapter are intended to draw attention to the fact that practitioners must carefully consider what kind of open innovation strategy is right for their business and intended objective. Not all modes of inbound open innovation provide the same returns depending on the nature of innovation in question, and overzealous external sourcing can actually lead to inferior returns and performance. Instead, focused efforts are likely to bring optimal results. The documented benefits of openness aside, history provides us with no shortage of *management fads* (or the more politically correct term of *management fashions*) whereby firms rush to adopt the latest and greatest management practices in hopes of gaining competitive advantage or even for fear of not being viewed as progressive (Phillips Carson, Lanier, Carson, & Guidry, 2000). Open innovation's rapid growth shares many characteristics of a management fashion, and there is danger that it will be unsustainable if the open innovation process is not properly managed and approached with realistic expectations (Gassmann, et al., 2010: 213). Often the expectations on managers to adopt the latest "big thing" in management, in this case open innovation, result in pressure to implement a process which may or may not be the most rational, effective course of action. One study found that firms employing popular management techniques did not have higher economic performance – although they were perceived as more admired and competent (Straw & Epstein, 2000). Clearly firms should engage in open innovation practices. However, as with any emerging management practice, there is danger of managers jumping in with both feet first, and after the realities of the difficulty of implementing any new process set in, abandoning it. The implication for firms is to embrace an open innovation strategy with selective external sources and particular objectives in mind while maintaining investments in internal capabilities.

In addition to implications for management, policy makers are now increasingly considering open innovation (OECD, 2008: 113-127). Some of the means in which policy is directed towards fostering open innovation include (1) technology and knowledge transfer policies; (2) the organization of the public sector research base such as universities and science parks; (3) establishment of public innovation intermediaries (Lee, Park, Yoon, & Park, 2010); (4) public funding for open innovation research and projects with a strong collaborative nature; and (5) policies directed towards the fair and tradable use of IP, i.e. the appropriation regime. The implications of this research for policy makers are twofold. First, open innovation clearly creates value but brings with it yet unknown difficulties. Knowledge is often "sticky" and difficult to transfer or value (von Hippel, 1988), especially for firms with few internal capabilities (Cohen & Levinthal, 1989). That is, investment in resources and capabilities internal to the firm still matter, both for the creation of new knowledge, whether tradable or destined for internal use, and for the generation of the absorptive capacity required to recognize and commercialize existing external knowledge. Open innovation is not a panacea, so policy makers must carefully consider policies which divert public funding away from the development of internal firm R&D to schemes simply supporting open innovation. Second, there is likely not a viable "one-policy-fits-all" approach to facilitating open innovation practices among firms. As argued in this chapter, firms are very heterogeneous in how they source and apply external knowledge, and furthermore the relevance of inbound open innovation may vary according to any number of factors such as a firm's industry and whether product or process innovation is the intended outcome. Policies directed at fostering open innovation will likely need to consider how the intended outcome may be contingent upon the idiosyncrasies of various industries and types of innovation.

ACKNOWLEDGMENT

The foundation for this work was sparked during a research stay gratefully sponsored by Keld Laursen in the Department of Innovation and Organizational Economics at Copenhagen Business School, Denmark, made possible by a researcher mobility grant from the PRIME Network. The econometric techniques benefited from coursework during a studentship at London Business School, funded by a mobility grant from the Spanish Scientific Research Council (CSIC), and a research stay at SKEMA Business School in Sophia Antipolis, France. I also thank my advisor, Jordi Molas-Gallart, for comments, support, and guidance.

REFERENCES

Anderson, B., Howells, J., Hull, R., Miles, I., & Roberts, J. (Eds.). (2000). *Knowledge and innovation in the new service economy*. Cheltenham: Edward Elgar Publishing.

Anderson-Macdonald, S., & Kask, S. (2010). *Service Innovation.* Marketing Science Institute: No. 10-304, Conference Summary, 21-23 September, London Business School.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Baetjer, H. (2000). Capital as Embodied Knowledge: Some Implications for the Theory of Economic Growth. *The Review of Austrian Economics*, *13*(2), 147–174. doi:10.1023/A:1007808618703

Barras, R. (1986). Towards a theory of innovation in services. *Research Policy*, *15*(4), 161–173. doi:10.1016/0048-7333(86)90012-0

Barras, R. (1990). Interactive innovation in financial and business services: The vanguard of the service revolution. *Research Policy*, *19*(3), 215–237. doi:10.1016/0048-7333(90)90037-7

Baumol, W. J. (1967). Macroeconomics of unbalanced growth: the anatomy of urban crisis. *The American Economic Review*, *57*(3), 415–426.

Berry, L. L., Venkatesh, S., Parish, J., Cadwallader, S., & Dotzel, T. (2006). Creating New Markets Through Service Innovation. *MIT Sloan Management Review*, *47*(2), 56–63.

Boden, M., & Miles, I. (Eds.). (2000). *Services and the Knowledge-Based Economy*. London: Continuum.

Brown, K. A. (1988). *Inventors at Work: Interviews with 16 Notable American Inventors*. Microsoft Press.

Cameron, C. A., & Trivedi, P. K. (2010). *Microeconomics Using Stata*. College Station, TX: Stata Press.

Chandy, R., Hopstaken, B., Narasimhan, O., & Prabhu, J. (2006). From Invention to Innovation: Conversion Ability in Product Development. *JMR, Journal of Marketing Research*, *43*(3), 494–508. doi:10.1509/jmkr.43.3.494

Chesbrough, H. (2003a). The logic of open innovation: Managing Intellectual Property. *California Management Review*, *45*(3), 33–58.

Chesbrough, H. (2003b). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.

Chesbrough, H. (2006). *Open Business Models: How to Thrive in the New Innovation Landscape*. Boston: Harvard Business School Press.

Chesbrough, H., & Crowther, A. K. (2006). Beyond high tech: early adopters of open innovation in other industries. *R & D Management*, *36*(3), 229–236. doi:10.1111/j.1467-9310.2006.00428.x

Chesbrough, H., & Garman, A. R. (2009). How Open Innovation Can Help You Cope in Lean Times. *Harvard Business Review*, *87*(12), 68–76.

Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, *11*(3), 529–555. doi:10.1093/icc/11.3.529

Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.). (2006). *Open Innovation: Researching a New Paradigm*. New York: Oxford University Press.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, *28*(3), 591–605. doi:10.2307/1910133

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and Learning: The Two Faces of R & D. *The Economic Journal*, *99*(397), 569–596. doi:10.2307/2233763

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive-Capacity - A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152. doi:10.2307/2393553

Dahlander, L., & Gann, D. M. (2010). How open is innovation? *Research Policy*, *39*(6), 699–709. doi:10.1016/j.respol.2010.01.013

Drakos, N. (2008). Case Study: How Cisco Embraced Crowdsourcing with Its I-Prize Initiative (pp. 1-6). Gartner Report G00156482.

Drejer, I. (2004). Identifying innovation in surveys of services: a Schumpeterian perspective. *Research Policy*, *33*(3), 551–562. doi:10.1016/j.respol.2003.07.004

Edquist, C., Hommen, L., & McKelvey, M. (2001). *Innovation and Employment: Process versus Product Innovation*. Cheltenham: Edward Edgar Publishing.

Flowers, S. (2007). Organizational capabilities and technology acquisition: why firms know less than they buy. *Industrial and Corporate Change*, *16*(3), 317–346. doi:10.1093/icc/dtm011

Fredberg, T., Elmquist, M., & Ollila, S. (2008). Managing Open Innovation: Present Findings and Future Directions (pp. 1-64). VINNOVA - Verket foer Innovationssystem/Swedish Governmental Agency for Innovation Systems.

Frenz, M., Girardone, C., & Ietto-Gillies, G. (2005). Multinationality Matters in Innovation: The Case of the UK Financial Services. *Industry and Innovation*, *12*(1), 65–92. doi:10.1080/1366271042000339067

Gallouj, F. (1998). Innovating in reverse: services and the reverse product cycle. *European Journal of Innovation Management*, *1*(3), 123–138. doi:10.1108/14601069810230207

Gallouj, F., & Djellal, F. (Eds.). (2010). *The handbook of innovation and services: a multi-disciplinary perspective*. Edward Elgar.

Gallouj, F., & Savona, M. (2009). Innovation in services: a review of the debate and a research agenda. *Journal of Evolutionary Economics*, *19*(2), 149–172. doi:10.1007/s00191-008-0126-4

Gassmann, O., Enkel, E., & Chesbrough, H. (2010). The future of open innovation. *R & D Management*, *40*(3), 213–221. doi:10.1111/j.1467-9310.2010.00605.x

Gourieroux, C., Montfort, A., & Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica*, *53*(3), 701–720. doi:10.2307/1913472

Hausman, J., Hall, B., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica*, *52*(4), 909–938. doi:10.2307/1911191

Hirsch-Kreinsen, H. (2008). "Low-Tech" Innovations. *Industry and Innovation*, *15*(1), 19–43. doi:10.1080/13662710701850691

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*(4), 331–343. doi:10.1002/smj.582

Huston, L., & Sakkab, N. (2006). Connect and develop: Inside Procter & Gamble's new model for innovation. *Harvard Business Review*, *84*(3), 58–66.

Katila, R. (2002). New Product Search over Time: Past Ideas in Their Prime? *Academy of Management Journal*, *45*(5), 995–1010. doi:10.2307/3069326

Katila, R., & Ahuja, G. (2002). Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction. *Academy of Management Journal*, *45*(6), 1183–1194. doi:10.2307/3069433

Katz, R., & Allen, T. J. (1982). Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R&D Project Groups. *R & D Management*, *12*(1), 7–20. doi:10.1111/j.1467-9310.1982.tb00478.x

Kennedy, P. (2003). *A Guide to Econometrics* (5th ed.).

Kessler, E. H., Bierly, P. E., & Gopalakrishnan, S. (2000). Internal vs. external learning in new product development: effects on speed, costs and competitive advantage. *R & D Management*, *30*(3), 213–224. doi:10.1111/1467-9310.00172

Kettinger, W. J., & Grover, V. (1995). Special section: toward a theory of business process change management. *Journal of Management Information Systems*, *12*(1), 9–30.

Koput, K. W. (1997). A chaotic model of innovative search: some answers, many questions. *Organization Science*, *8*(5), 528–542. doi:10.1287/orsc.8.5.528

Laden, K. (1996). `Not invented there,' or, the other person's dessert always looks better! *Research Technology Management*, *39*(6), 10.

Laursen, K., & Salter, A. (2004). Searching high and low: what types of firms use universities as a source of innovation? *Research Policy*, *33*(8), 1201–1215. doi:10.1016/j.respol.2004.07.004

Laursen, K., & Salter, A. (2006). Open for innovation: The role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal*, *27*(2), 131–150. doi:10.1002/smj.507

Lee, S., Park, G., Yoon, B., & Park, J. (2010). Open innovation in SMEs - An intermediated network model. *Research Policy*, *39*(2), 290–300. doi:10.1016/j.respol.2009.12.009

Leiponen, A., & Helfat, C. E. (2010). Innovation objectives, knowledge sources, and the benefits of breadth. *Strategic Management Journal*, *31*(2), 224–236.

Levinthal, D. A., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior & Organization*, *2*(4), 307–333. doi:10.1016/0167-2681(81)90012-3

Lichtenthaler, U. (2011). Open Innovation: Past Research, Current Debates, and Future Directions. *academy of Management Perspectives, 25*(1).

Lichtenthaler, U., & Ernst, H. (2006). Attitudes to externally organising knowledge management tasks: a review, reconsideration and extension of the NIH syndrome. *R & D Management*, *36*(4), 367–386. doi:10.1111/j.1467-9310.2006.00443.x

Lowe, J., & Taylor, P. (1998). R&D and technology purchase through licence agreements: complementary strategies and complementary assets. *R & D Management*, *28*(4), 263–278. doi:10.1111/1467-9310.00103

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. doi:10.1287/orsc.2.1.71

Menon, T., & Pfeffer, J. (2003). Valuing Internal vs. External Knowledge: Explaining the Preference for Outsiders. *Management Science*, *49*, 497–513. doi:10.1287/mnsc.49.4.497.14422

Mesaglio, M., & Hunter, R. (2008). *Mastering the Paradoxes of Innovation*. CIO Signature.

Nelson, R. R., & Winter, S. G. (1973). Toward an Evolutionary Theory of Economic Capabilities. *The American Economic Review*, *63*(2), 440–449.

Niehaves, B. (2010). Open process innovation: The impact of personnel resource scarcity on the involvement of customers and consultants in public sector BPM. *Business Process Management Journal*, *16*(3), 377–393. doi:10.1108/14637151011049412

Ocasio, W. (1997). Towards an Attention-Based Theory of the Firm. *Strategic Management Journal*, *18*(S1), 187–206. doi:10.1002/(SICI)1097-0266(199707)18:1+<187::AID-SMJ936>3.3.CO;2-B

OECD. (2000). *The Service Economy* (pp. 1–52). Paris: Business and Industry Policy Forum Series.

OECD. (2008). *Open Innovation in Global Networks*. Paris: OECD Publications.

Phillips Carson, P., Lanier, P. A., Carson, K., & Guidry, B. N. (2000). Clearing a Path through the Management Fashion Jungle: Some Preliminary Trailblazing. *Academy of Management Journal*, *43*(6), 1143–1158. doi:10.2307/1556342

Pine, J. B., & Gilmore, J. H. (1999). *The Experience Economy: Work is Theatre and Every Business a Stage*. Boston: Harvard Business Press.

Pisano, G. P. (1997). *The Development Factory: Unlocking the Potential of Process Innovation*. Boston: Harvard Business School Press.

Reichstein, T., & Salter, A. (2006). Investigating the sources of process innovation among UK manufacturing firms. *Industrial and Corporate Change*, *15*(4), 653–682. doi:10.1093/icc/dtl014

Reitzig, M., & Wagner, S. (2010). The Hidden Costs of Outsourcing: Evidence from Patent Data. *Strategic Management Journal*, *31*(11), 1183–1201. doi:10.1002/smj.852

Rohrbeck, R., Hölzle, K., & Gemünden, H. G. (2009). Opening up for competitive advantage: How Deutsche Telekom creates an open innovation ecosystem. *R & D Management*, *39*(4), 420–430. doi:10.1111/j.1467-9310.2009.00568.x

Rouvinen, P. (2002). Characteristics of product and process innovators: some evidence from the Finnish innovation survey. [Article]. *Applied Economics Letters*, *9*(9), 575–580. doi:10.1080/13504850110108102

Rubalcaba, L. (2007). *The new service economy: Challenges and policy implications for Europe*. Chentelham, UK: Edward Elgar Publishing.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, *88*(4), 641–658. doi:10.1162/rest.88.4.641

Simon, H. A. (1997). *Administrative Behavior* (4th ed.). New York: The Free Press.

Sirilli, G., & Evangelista, R. (1998). Technological innovation in services and manufacturing: results from Italian surveys. *Research Policy*, *27*(9), 881–899. doi:10.1016/S0048-7333(98)00084-5

Spithoven, A., Clarysse, B., & Knockaert, M. (2010). Building absorptive capacity to organise inbound open innovation in traditional industries. *Technovation*, *30*(2), 130–141. doi:10.1016/j.technovation.2009.08.004

Straw, B. M., & Epstein, L. D. (2000). What Bandwagons Bring: Effects of Popular Management Techniques on Corporate Performance, Reputation, and CEO Pay. *Administrative Science Quarterly*, *45*(3), 523–556. doi:10.2307/2667108

Tether, B. S., Howells, J., Bessant, J., Davies, A., Voss, C., Zomerdijk, L., et al. (2007). Innovation in Services. In D. E. Papers (Ed.), *DTI Occasional Paper NO. 9* (pp. 198).

Tsai, K.-H., & Wang, J.-C. (2007). Inward technology licensing and firm performance: a longitudinal study. *R & D Management*, *37*(2), 151–160. doi:10.1111/j.1467-9310.2007.00464.x

Tsai, M.-H., & Wang, J.-C. (2009). External technology sourcing and innovation performance in LMT sectors: An analysis based on the Taiwanese Technological Innovation Survey. *Research Policy*, *38*(3), 518–526. doi:10.1016/j.respol.2008.10.007

Van de Vrande, V., Vanhaverbeke, W., & Gassmann, O. (2010). Broadening the scope of open innovation: past research, current state and future directions. *International Journal of Technology Management*, *52*(3-4), 221–235. doi:10.1504/IJTM.2010.035974

Vega-Jurado, J., Gutierrez-Gracia, A., & Fernandez-de-Lucio, I. (2009). Does external knowledge sourcing matter for innovation? Evidence from the Spanish manufacturing industry. *Industrial and Corporate Change*, *18*(4), 637–670. doi:10.1093/icc/dtp023

Vivero, R. L. (2002). The impact of process innovations on firm's productivity growth: the case of Spain. [Article]. *Applied Economics*, *34*(8), 1007–1016. doi:10.1080/00036840010019684

von Hippel, E. (1988). *The Sources of Innovation*. Oxford: Oxford University Press.

West, J., Vanhaverbeke, W., & Chesbrough, H. W. (2006). Open Innovation: A Research Agenda. In H. W. Chesbrough, W. Vanhaverbeke & J. West (Eds.), *Open Innovation: Researching a New Paradigm* (pp. 285-307). New York: Oxford University Press.

Yadav, M. S., Prabhu, J., & Chandy, R. (2007). Managing the Future: CEO Attention and Innovation Outcomes. *Journal of Marketing*, *71*(4), 84–101. doi:10.1509/jmkg.71.4.84

ADDITIONAL READING

Acha, V. (2008). *Open by Design: The Role of Design in Open Innovation* (Report, D. R., Ed.). London: Department for Innovation, Universities and Skills.

Alexy, O., Criscuolo, P., & Salter, A. (2009). Does IP Strategy Have to Cripple Open Innovation? *MIT Sloan Management Review*, *51*(1), 71–77.

Amara, N., & Landry, R. (2005). Sources of information as determinants of novelty of innovation in manufacturing firms: evidence from the 1999 statistics Canada innovation survey. *Technovation*, *25*(3), 245–259. doi:10.1016/S0166-4972(03)00113-5

Asakawa, K., Nakamura, H., & Sawada, N. (2010). Firms' open innovation policies, laboratories' external collaborations, and laboratories' R&D performance. *R & D Management*, *40*(2), 109–123. doi:10.1111/j.1467-9310.2010.00598.x

Boudreau, K., &Lakhani, K. R. (2009). How to Manage Outside Innovation. *MIT Sloan Management Review, Summer*, 69-76.

Boudreau, K. (2010). Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. *Management Science*, *56*(10), 1849–1872. doi:10.1287/mnsc.1100.1215

Dodgson, M., Gann, D., & Salter, A. (2008). *The Management of Technological Innovation*. Oxford: Oxford University Press.

Drossau, O., Krempl, S., & Poltermann, A. (Eds.). (2006). *Die wunderbare Wissensvermehrung: Wie Open Innovation unsere Welt revolutioniert*. Hannover: Heise Zeitschriften Verlag GmbH.

Enkel, E., & Gassmann, O. (2010). Creative imitation: exploring the case of cross-industry innovation. *R & D Management*, *40*(3), 256–270. doi:10.1111/j.1467-9310.2010.00591.x

Enkel, E., Gassmann, O., & Chesbrough, H. (2009). Open R&D and open innovation: exploring the phenomenon. *R & D Management*, *39*(4), 311–316. doi:10.1111/j.1467-9310.2009.00570.x

Fasnacht, D. (2009). *Open Innovation in the Financial Services: Growing Through Openness, Flexibility, and Customer Integration*. Berlin: Springer-Verlag.

Gassmann, O. (2006). Opening up the innovation process: towards an agenda. *R & D Management*, *36*(3), 223–228. doi:10.1111/j.1467-9310.2006.00437.x

Harhoff, D., Henkel, J., & von Hippel, E. (2003). Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research Policy*, *32*(10), 1753–1769. doi:10.1016/S0048-7333(03)00061-1

Howe, J. (2009). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. New York: Random House.

Huston, L., & Sakkab, N. (2007). Implementing Open Innovation. *Research Technology Management*, *50*(2), 21–25.

Koput, K. W. (1997). A chaotic model of innovative search: some answers, many questions. *Organization Science*, *8*(5), 528–542. doi:10.1287/orsc.8.5.528

Lane, P. J., Koka, B. R., & Pathak, S. (2006). The Reification of Absorptive Capacity: A Critical Review and Rejuvenation of the Construct. [Article]. *Academy of Management Review*, *31*(4), 833–863. doi:10.5465/AMR.2006.22527456

Leonard-Barton, D. (1995). *Wellsprings of knowledge: building and sustaining the sources of innovation*. Harvard Business School Press.

Michel, S., Brown, S. W., & Gallan, A. S. (2008). Service-Logic Innovations: How to Innovate Consumers, Not Products. *California Management Review*, *50*(3), 49–65.

Tao, J., & Magnotta, V. (2006). How Air Products and Chemicals "identifies and accelerates". *Research Technology Management*, *49*(5), 12–18.

Thomke, S. (2003). R&D Comes to Services: Bank of America's Pathbreaking Experiments. *Harvard Business Review*, *81*(4), 70–79.

Vega-Jurado, J., Guiterrez-Gracia, A., & Fernandez de Lucio, I. (2008). Analyzing the determinants of firm's absorptive capacity: beyond R&D. *R & D Management*, *38*(4), 392–405. doi:10.1111/j.1467-9310.2008.00525.x

Veugelers, R., & Cassiman, B. (1999). Make and buy in innovation strategies: evidence from Belgian manufacturing firms. *Research Policy*, *28*(1), 63–80. doi:10.1016/S0048-7333(98)00106-1

von Hippel, E. (2005). *Democratizing Innovation*. Cambridge, MA: The MIT Press.

Zahra, S. A., & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. *Academy of Management Review*, *27*, 185–203.

This work was previously published in Open Innovation in Firms and Public Administrations: Technologies for Value Creation, edited by Carmen de Pablos Heredero and David López, pp. 15-36, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 11
Production Competence and Knowledge Generation for Technology Transfer:
A Comparison between UK and South African Case Studies

Ian Hipkin
École Supérieure de Commerce de Pau, France

EXECUTIVE SUMMARY

The ability to execute the physical part of manufacturing will assume greater importance as new technology and knowledge become significant drivers of strategic direction. The case studies described in this chapter address the interaction between technology transfers (TT), production competence and knowledge in enhancing performance in manufacturing organizations. Reference to British and South African case studies provides a useful comparison of production competence in the developed and developing world. In both countries, operators and maintainers lacked detailed knowledge of equipment functionality and performance parameters. United Kingdom (UK) companies demonstrated a deeper understanding of fundamental principles of the underlying production process, enabling them to remedy production deficiencies more thoroughly. South African companies showed greater management commitment to training and group solving approaches.

ORGANIZATION BACKGROUND

New equipment and systems play a vital role as production-oriented firms strive to meet changing customer demands and increased competition. Technology is key to achieving competitive advantage to this and requires an exchange of technical know-how and the transfer of competencies to facilitate organizational learning, consolidation of new technologies, and establishing routines and production skills (Lynskey, 1999). Frequently recurring factors that contribute to poor performance of new technology in both developed and

DOI: 10.4018/978-1-4666-1945-6.ch011

developing countries are inadequate knowledge of plant and processes, managerial commitment, information systems, and lack of production competence. This chapter describes an investigation into the impact of such factors on TT in a number of UK and South African manufacturing organizations that have recently acquired new technology, and assesses how similarities and differences in these organizations affect production performance.

SETTING THE STAGE

It has long been suggested that equipment performance is dependent on production competence (Ferdows and De Meyer, 1990; Leonard-Barton, 1995). Employees who understand how something is made are better able to manage manufacturing practices than those who deal only with the underlying symptoms of a process (Gourley, 2006). An environment where knowledge assimilation and sharing generate continuous learning capability lends itself to absorptive capacity (Tu et al, 2006). Early proponents such as Cohen and Levinthal (1990) see absorptive capacity (AC) as the ability of a firm to recognise and utilise the value of new information that comes from continuous learning. Mowery and Oxley (1995) construe AC as a comprehensive set of skills that engages with the tacit component of transferred and imported knowledge. Jones and Craven (2001) emphasise the need for communication networks that diffuse knowledge and technology.

Production competence and knowledge generation may be seen as a function of AC, which Zahra and George (2002) define as: “AC is a set of organizational routines and processes by which firms acquire, assimilate, transform and exploit knowledge to produce a dynamic organizational capability” (p.185). Like many commentators these authors view AC as an essentially organizational phenomenon.

In order to understand a knowledge-based process Rouse and Daellenbach (2002) suggest that we should follow the value generation trail back to its source. This requires “unpacking the process, beginning with performance, then looking for sources of advantage in the form of capabilities and competencies. Unless the sources of firm-specific superiority can be understood, the nature of the competitive advantage is doomed to remain largely a mystery. We will merely have substituted a black box of competencies for the black box of organizations.” (p.966) Rouse and Daellenbach (2002) also suggest a theoretical and empirical understanding of what occurs in TT as well as how this takes place:

- Resources, technology and knowledge (tangible and intangible) are bundled, linked and incorporated as management processes.
- Management processes and structural relationships are converted and organised into routines and systems.
- Routines and systems formalise capabilities that lead to production competencies.
- Competencies become the means whereby products and services are generated to provide value and competitive advantage.

Resources and Knowledge

Resource-based paradigms help to explain the success of firms through an analysis of technological skills, complementary assets and routines. Capabilities are clusters of tangible and intangible abilities that span individuals and groups (Coates and McDermott, 2002). Manufacturing firms need a competitive focus to optimise quality, cost and flexibility and develop competencies in the long term, so it is necessary to identify the tangible and intangible knowledge and skills required, and formulate policies to prioritise these. Knowledge further assists designers in understanding the consequences of any changes that may be made in improving product features (Cristofolini et al, 2009).

Knowledge is a useful lens through which to view the development of competencies. Interlinking technical areas allows firms to combine multiple knowledge and functional domains to overcome challenges in implementing new technology. Pools of knowledge lead to competencies so an essential factor in their emergence is an orientation towards learning and developing knowledge. Training and familiarisation are prerequisites for the creation and use of new competencies. Knowledge is a resource that supports a firm's ability to excel in achieving these competitive priorities. Among the reasons for inadequate knowledge is a firm's over-optimism about the knowledge of its own processes, and its inability to build, debug and operate new processes (Bohn, 1994).

Management Support

Firms need management processes that will transcend organizational boundaries and co-ordinate activities across diverse environments and social dynamics (Dekkers, 2009). They benefit from leadership roles played by proactive individuals, boundary spanners and technology champions (Irwin et al, 1998). These facilitate crossing technical boundaries to find multidisciplinary solutions through innovation and bringing together diverse experts. Managers should lead in offering support for new equipment, while being sufficiently flexible to accommodate the emerging behaviour and actions that unfold from operators of the technology. This requires rethinking established processes that affect, or are influenced by, the technology.

Technological competence depends on the ability and contribution of development teams, but managers set the process in motion, provide facilities for training and encourage organizational learning as a means to develop requisite capabilities. Managers should also encourage the process of knowledge acquisition, through "legitimizing familiarisation activities" (Dimnik and Johnston, 1993, p.161). Crossan and Berdrow (2003) emphasise the development of shared understanding amongst individuals and taking coordinated action through mutual adjustment. This will initially be unstructured and informal, but may be institutionalised by embedding the learning that has occurred by into systems, structures and procedures. Within a constantly changing environment, managers must manage the tension between institutionalised learning and the processes of intuitive interpretation and integration.

The more individuals work with colleagues and benefit from physical and psychological proximity, the more they develop mutual trust and are willing to share their specialist knowledge (Uzzi and Lancaster, 2003). The effect is faster knowledge assimilation through observation, the facility to draw on a pool of contextual knowledge, and a shared sense of where expertise lies. Reagans et al (2005) refer to this as a 'relationship-specific' activity where knowledge is assimilated and embedded in the link connecting people performing distinct roles.

Information and Systems

Management scientists and operations researchers have developed decision-making tools for operations and maintenance that are intended to replace subjective decisions with objective ones and assist managerial control through faster, more comprehensive and accurate knowledge of operations, and reducing idiosyncratic judgement and mystique. However, Pintelon and Gelders (1992) claim that little use is made of such techniques and challenge the benefits of a large quantity of explicit data, even if it can be gathered. Researchers can identify some key information systems when assessing TT, but cannot access others that are invisible and buried in the minds of the actors. For example, computerised maintenance management systems have been developed as a form of data repository. Although maintainers perceive the lack of failure data as a barrier to maintenance intervention, it is questionable how much useful data is stored in computerised maintenance systems. Aharoni

(1993) suggests that the diversity of information systems emphasises how the richness in quality and nature of information systems can be imbued with value, context and meaning. However, researchers must assess whether information systems enhance production or support organizational governance through a form of indirect exertion of authority. Even if the systems governing each individual's task within an organizational routine are explicit, the complex totality of the system may well be unknown to most participants, and is therefore tacit. Few systems can accommodate knowledge, know-how, social capital, and other socially complex, difficult-to-transfer data.

Production Competence

Coates and McDermott (2002) define competence as: "a bundle of aptitudes, skills, and technologies that the firm performs better than its competitors." (p.443) The importance of understanding equipment performance is widely supported in the literature on process competencies and organizational learning (Leonard-Barton, 1995; Marmier et al, 2009). Processes dependent on technological subsystems require subtle skills and expertise that are often difficult to codify. The secrets and capabilities of a complex process are understood most effectively through direct, first-hand observation and interaction (Lynskey, 1999).

Galbreath (2005) posits that knowledge is assimilated as capability when familiarity with tangible assets through employee learning, and interaction between management and operational personnel become evident in operational routines. Training can at best only go part-way in imparting this expertise. Assimilation is delayed in conditions of uncertainty when knowledge elements lie beyond an organization's sphere of expertise (Rosenkopf and Nerkar, 2001). This may be due to novelty and layers of complexity which even knowledgeable operators are unable to comprehend (Dehane et al, 2008; Sorenson et al, 2006).

Macher (2006) notes that problem solving is often a trial-and-error exercise where particular parameters are modified, learned about, and improved. A range of solutions is likely to emerge, and their impact may be interpreted in differing ways by multiple stakeholders with possibly contradictory interests. Above all, the success of a technology depends to a large degree on the competencies of personnel, since documented and accessible explicit knowledge may form only a small part of the knowledge that should be shared. The management of collective knowledge in firms and the ability of individuals to work together to convert knowledge into customer value are essential competencies. The challenge for researchers is to develop theory that will "identify unique, proprietary, and special abilities in manufacturing firms and understand how competencies that lead to competitive advantage both occur and are capitalized upon in the firm" (Coates and McDermott, 2002, p.438).

CASE DESCRIPTION

Four organizations that had recently implemented new technology were selected for this study. Broadly comparable firms in two sectors were studied (paper and food): two from the United Kingdom (referred to as PaperUK and FoodUK) and two from South Africa (PaperSA and FoodSA). A brief description of each is provided.

- *PaperUK*: This 30 year old UK paper operated ageing equipment, beset with breakdowns. It produced speciality paper for the printing industry and operated several paper machines, coaters and a finishing department. Production targets were only met with excessive maintenance overtime. Low profitability meant that capital expenditure had been severely limited, so the life of existing equipment was to be extended through maintenance and some small technological advancements. Management had installed a computerised maintenance management system (CMMS), but this was recognised as deficient in its lack of user friendliness. This could not be construed as a knowledge-based operation as the vast quantities of data accumulated over the years were difficult to access, and had little credibility. No direct interface between the CMMS and the production system existed. As part of a long-term plant upgrade a supplier was appointed to install a new coater on one paper machine in close association with a team of paper mill staff from engineering, operations and maintenance. Commissioning of the new coater was accompanied by the installation of a new CMMS, in order to raise the level of knowledge of plant operation and failure data. This represented a substantial upgrade potential for the maintenance function. Reliable information was seen as essential for effective maintenance. The CMMS was to be used to attain the original functionality of equipment, and to ensure higher plant availability. 12 individuals were interviewed in PaperUK.

- *FoodUK*: This UK food company was confronted by tremendous cost competition from global players, and was attempting to find niche markets to build on its quality competitive advantage. Lower operating costs were essential provided quality standards could be maintained. Labour intensive quality checking was no longer sustainable: this was expensive and not sufficiently reliable. Customers had been criticising the company for not meeting its agreed acceptable quality level standards, and several major customers had been lost. In some areas production stoppages could be tolerated depending on downline processes and orders. Opportunistic repairs during production stoppages were becoming the dominant way of doing maintenance, although it was recognised that this was not effective maintenance. A large backlog of data had not been entered into the existing maintenance management system. This food factory installed laser screening equipment to monitor raw materials used in the manufacture of tinned foods and take remedial action to prevent substandard products from entering the manufacturing process. The challenge was to ensure that operators were sufficiently knowledgeable and proficient in utilising the new technology, and to guarantee that strict quality standards were met. 13 individuals were interviewed in Food UK.
- *PaperSA*: The strategy of this South African paper mill was to produce high quality paper at a cost lower than its competitors. Despite an industry characterised by depressed prices, there were profitable markets for the company's products. The company had no knowledge-based system for maintenance since the few historical failure records were based on perceptions of certain members of the engineering and maintenance departments, as well as opportunistic maintenance which was undertaken during scheduled and non-scheduled production stoppages. Maintenance and production staff were required to enter failure and performance data at terminals in the operating control room. This paper mill comprised paper machines, winders, rewinders, core and log handling, and packing machines. Various new projects had been undertaken including an upgrade of the logsaw and associated equipment, and a new packaging machine. The choice of equipment and quality standards was made on the basis of extensive consumer market research which indicated where customer values lay. Significant management was expended in ensuring that operating staff

accepted the new quality standards, which were at times at variance with previous operating practices. 9 individuals were interviewed in PaperSA.

- *FoodSA*: This food-processing factory had pockets of low unit cost competitive advantage brought about by earlier investment in new technologies. The low cost advantage was eroded by fierce competition, and extensive down time on several sorting machines. Maintenance had been predominantly preventive, but management felt that a more structured approach was required, which required information to be recorded in a computerised system. Production departments were to be convinced that maintainers possessed valuable knowledge that would reduce the frequency and duration of production stoppages. Although a number of variables in raw materials influenced the quality of the end product, the level of output variance was limited as quality was strictly in accordance with American fast-food quality. New sorting equipment had been acquired, but no maintenance history was available. The challenge for the company was to acquire sufficient knowledge of the equipment to ensure high levels of available and reliability. 10 individuals were interviewed in FoodSA.

CURRENT CHALLENGES FACING THE ORGANIZATIONS

This section recounts occurrences and experiences in the case organizations in relation to the main points identified in the 'setting the scene' section of this chapter. One of the common features in all case studies was that previous operating practices and equipment were not delivering the required performance. New technology was introduced in each case, but knowledge of new processes was lacking, and managers were not sure how this deficiency should be addressed. Their actions were essentially trial and error, and largely unstructured.

Resources and Knowledge

Resources are the assets owned or controlled by a firm, and include new equipment and knowledge vested in people. In investigating the challenges facing the case organizations, it is instructive to identify subtle knowledge acquisition, working practices and adjustments that maximise the benefits that can be derived from new technology.

New laser equipment at FoodUK replaced manual raw vegetable quality monitoring. Settings by suppliers and plant engineers were based on trial and error. In assessing the quality of the incoming product and monitoring the output, operators were able to use their previous experience of the manual process (tacit knowledge) to set the machine to ensure minimum rejection of acceptable products and maximum rejection of unacceptable products. In order to achieve this, a great deal of management time and effort was dedicated to persuading operators of the value of this knowledge. This was necessary since operators had initially distanced themselves from the technology, expressing the sentiment that the technology had effectively displaced them.

FoodSA also had varying qualities of incoming raw material. Machines were set to allow sizes to vary between limits, depending on customer requirements. Operators followed procedures laid down by supervisors since they were not aware of operating parameters within which outsized materials could be tolerated. FoodSA's explicit instructions were not relaxed to accommodate the type of tacit knowledge demonstrated in FoodUK.

These examples illustrate the importance of delicate adjustments that lie beyond standard operating procedures. According to one respondent at FoodUK the basis for successful use of this equipment resided in 'an uncanny knack' by certain operators to apply incremental fixes and fine adjustments to meet specified quality requirements. Despite their documentation and engineering drawings (explicit knowledge), process engineers were unable to do this. FoodSA's procedures did not encourage the degree of flexibility that would permit some experimentation on the basis of tacit knowledge.

Personnel in the UK companies possessed greater knowledge than their South African counterparts. Although only some events are recounted, it was clear that the experience and initiatives of the UK operators in particular (on the basis of tacit knowledge) added significantly to production output. In neither food company was there any formal approach to knowledge acquisition. PaperSA did make a concerted effort to acquire knowledge of customer quality standards, and to ensure that products adhered to these requirements.

Management Support

Managers in the case studies acknowledged the importance of training staff members involved with operating and maintaining new equipment. PaperUK engaged the supplier of the coater equipment to train operators and maintenance craftsmen. However, the training did not encourage 'ownership' of the plant by the newly constituted team because several features of the coater replicated difficulties with older equipment and operators felt their opinions had not been taken into account when the new coater was specified. Management support was perceived to be wanting. The result was a lack of knowledge assimilation: while knowledge could be acquired through the training, operators did not accept or readily assimilate it.

Much discussion preceded the acquisition of the new equipment at PaperSA, especially regarding internal quality standards. It was possible to meet external requirements but internal standards remained a source of contention. The amount of glue sprayed onto a toilet roll to secure the final sheet (the 'tail') determined how many sheets were glued together while the length of 'tail' affected final packaging. Management established a team of operators, maintenance and process staff to address this type of issue, but only when a champion emerged in the group did team members demonstrate a concerted effort to investigate appropriate specifications for the equipment. In absorptive capacity terms, this reflected assimilation and transformation: the competitive advantage gained from first mover advantage in setting and adhering to appropriate standards was truly transforming.

In this instance the South African managers involved their staff more in discussions of the new technology. This provided a legitimate basis for the group to propose quantitative performance standards which were subsequently accepted as company quality policy, and led to an industry leadership position.

Information and Systems

As part of FoodSA's ISO9000 certification exercise, detailed production procedures were produced and systems established to provide an audit trail for quality inspections. Managers' attention was more directed towards compliance with the system than with production requirements.

PaperSA attempted to make use of its computerised system in determining maintenance intervention. However, existing schedules were done on a conservative time basis. Many failures were prevented, but no one could quantify the costs of such excessive maintenance. The maintenance system was thus of little use: some data existed, but seldom was this analysed to produce information, and no evidence was apparent of a knowledge based maintenance system.

PaperUK had a long-established computerised maintenance system, and managers were adamant that procedures should be based on failure history (data). Failures were recorded, but the data captured was not suitable for establishing maintenance tasks. The emphasis was on data input, rather than ensuring meaningful outputs.

Experiences in the case studies show that information systems were used to little effect. The nature of data collection was suspect, and procedures for analysing information were not satisfactory. While some data was thus collected, it was not adequately assimilated, and failed to transform the maintenance systems of any of the case studies. The findings of this study suggest that all case studies suffered from a lack of accurate data and inability or reluctance to use information systems: operators and maintainers were unable to base remedial action on any documented history and experience.

If time-based maintenance is adopted, then a period of time must elapse before meaningful data can be collected for new equipment. The case studies did not demonstrate that they were going to collect failure data for the new equipment in a meaningful manner, nor that this would be competently analysed.

Production Competence

In order to assess the levels of production competence in the four cases, respondents were asked to provide certain performance related information pertaining to their plant. The items contained in this paper related to the following (the results are summarised in Table 1):

- primary function of the new equipment, with performance standards in terms of output quantity and quality
- details of how non-achievement of the primary function would be detected, and the nature of the consequences
- details of protective systems (safety to humans and equipment protection)
- details of the most severe failures (in terms of cost or safety), with most appropriate remedial action, and the basis for this.

Table 1 shows, for example, that all respondents at FoodUK were able to describe the primary function of the laser technology, and the production supervisor and operator were able to quantify output. However, no one could provide precise quality standards.

The information in Table 1 and interview results confirm that production competence in the UK was greater than in South Africa, but they still demonstrate various degrees of ignorance about production and maintenance issues. While operators were generally familiar with the function and performance of their equipment, maintenance craftsmen were poorly informed. The implication is that those who maintain equipment were not fully aware of its functionality. While most respondents would identify the failure to fulfil the primary function, they were generally unaware of the consequences in the longer term (such as safety, environmental or health implications, costs, effects on customer service). There was a low level of awareness of safety and equipment protection. Respondents were familiar with the most common causes of failure, but none could define precisely the basis on which remedial action was taken (in terms of meeting functionality requirements, assessing root causes of failure, determining consequences of loss of functionality, maintenance tasks and intervals, and so on).

The topics discussed above provide some useful pointers pertaining to TT in the four cases. This section discusses implications of the findings for managers. Table 2 summarises the findings encountered in broad terms.

SOLUTIONS AND RECOMMENDATIONS

The research suggests that the way in which firms manage their resources and competencies will determine the extent to which these will result in enduring sources of competitive advantage through superior performance and functionality. The challenge for managers is to determine what distinctive ways and routines should be harnessed

Table 1. Summary of selected measures of production competence by company and position

FoodUK				
Measure of production competence	Prod supervisor	Operator	Craftsman	Process engineer
Details of primary function Output rate (quantity/time unit) Quantification of quality standards	Yes Yes No	Yes Yes No	Yes No No	Yes Yes Yes
Detection of failure of function Consequences of failure of function	Yes Yes	Yes No	Yes No	No Yes
Safety protective systems Equipment protection systems	Yes No	Yes No	Yes Yes	No Yes
Details of most severe failure Details of appropriate action Basis for remedial action	Yes Yes No	Yes Yes No	Yes Yes No	Yes No No
FoodSA				
Measure of production competence	Prod supervisor	Operator	Craftsman	Process engineer
Details of primary function: Output rate (quantity/time unit) Quantification of quality standards	Yes No No	Yes Yes No	No No No	Yes Yes Yes
Detection of failure of function Consequences of failure of function	Yes No	Yes No	Yes No	No No
Safety protective systems Equipment protection systems	Yes No	Yes No	No Yes	No Yes
Details of most severe failure Details of appropriate action Basis for remedial action	Yes No No	Yes No No	Yes Yes No	Yes No No
PaperUK				
Measure of production competence	Prod supervisor	Operator	Craftsman	Process engineer
Details of primary function: Output rate (quantity/time unit) Quantification of quality standards	Yes Yes Yes	Yes Yes No	Yes No No	Yes Yes Yes
Detection of failure of function Consequences of failure of function	Yes Yes	Yes Yes	Yes No	Yes Yes
Safety protective systems Equipment protection systems	No Yes	Yes Yes	Yes Yes	Yes No
Details of most severe failure Details of appropriate action Basis for remedial action	Yes Yes No	Yes Yes No	Yes Yes No	Yes Yes No
PaperSA				
Measure of production competence	Prod supervisor	Operator	Craftsman	Process engineer
Details of primary function: Output rate (quantity/time unit) Quantification of quality standards	Yes Yes No	Yes No No	Yes No No	Yes No Yes
Detection of failure of function Consequences of failure of function	Yes Yes	Yes No	Yes Yes	Yes Yes
Safety protective systems Equipment protection systems	No Yes	No No	Yes No	No No
Details of most severe failure Details of appropriate action Basis for remedial action	Yes No No	Yes No No	Yes Yes No	Yes No No

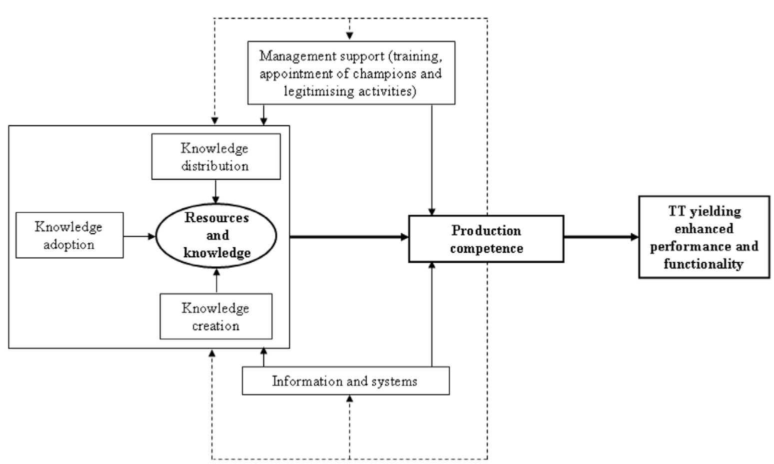
Table 2. Summary of findings

	FoodUK	PaperUK	FoodSA	PaperSA
1. Resources and knowledge	Tacit knowledge improved performance through an ability to apply subtle changes to process	Operator experience enabled minor adaptations to operating procedures to improve performance	Little scope for enhancing tacit knowledge because procedures were rigidly followed	Specified production processes prevented learning through experimentation. Little tacit knowledge evident
2. Management support	No special support evident to enhance TT	No special support evident to enhance TT	Training and consultation demonstrated management support	Extensive consultation enhanced TT through management support
3. Information and systems	Systems contributed little to TT	Systems contributed little to TT	Systems assisted in monitoring production output	Systems used incorrectly; excessive maintenance increased costs
4. Production competence	Reasonable levels of competence led to meeting desired performance standards in relatively short time	Reasonable operator and craftsman knowledge resulted in good flexibility to meet varying customer demands	Operators and craftsmen not familiar with plant and process. Large wastage of raw materials	Low levels of competence resulted in low initial output and variable quality

to achieve these. Findings show that the UK cases demonstrated greater knowledge and insight of processes, and operators and maintainers had higher levels of production competence than the South Africans. Management intervention in the South African cases was more supportive for TT. Information and systems were of little direct benefit in any of the cases. The transfer of knowledge to competencies is not a one-to-one linear process because of the indeterminate nature of knowledge and the diversity of competencies required for effective utilisation of new technology. Typical competencies include technological skills and complementary assets, as well as routines and capabilities that appear as clusters of tangible and intangible abilities spanning individuals and groups.

Competencies identify the key resources that need to be regenerated, expanded and built on in a firm's future activities, as tacit knowledge can generate more durable advantages (as was found in the UK cases). The feedback loops suggest that competencies help identify the tangible and intangible assets that need to be acquired, and

Figure 1. Relationship between resources, competence and performance



the direct attention that needs to be nurtured and prioritised inside the firm. This is illustrated in Figure 1.

Governance decisions and management support may have both beneficial and detrimental effects. A control spiral can result in yet further control (such as FoodSA's emphasis on procedures and documentation, rather than analysing the broader picture of what the true purpose is of such control). Incremental adjustments in a firm's governance structure may be desirable to permit evolutionary experimentation, encourage individual initiative and enhance organizational learning frameworks, through training, appointment of champions and legitimising activities.

The UK firms encouraged the creation and deployment of knowledge, and the findings suggest that knowledge as a resource accelerated technical progress. The primary motivation for encouraging the emergence and development of pools of new knowledge is therefore to achieve learning and competencies in technology, engineering and processes that will then enable the firm to pursue and develop new opportunities. Respondents in the cases found that technology transfer was frequently hindered by the tacitness of knowledge and complexity in new equipment (thus, the process engineer at FoodUK was unable to arrive at optimal settings).

This paper has investigated a number of issues relating to TT in two UK and two South African organizations. Of course, TT encompasses many more concepts than the four identified, and only limited generalisation from four case studies is possible. However, the study has demonstrated that a relationship certainly exists between knowledge, production competence and TT, as illustrated in Figure 1. This becomes apparent when the benefits of greater levels of knowledge are manifested in enhanced production competence, which in turn leads to improved performance and functionality. The UK companies benefited from the knowledge and competences vested within them, when compared to the South African firms. Yet, there is more to TT than knowledge and competences: management support and information systems can play an invaluable important role. Management support was stronger in the South African firms, but information systems did not play a significant part in any of the cases.

Several recommendations are appropriate for managers embarking on successful knowledge-based TT. Managers must develop resources and knowledge if full benefit is to derive from the technology. Management support is a necessary but not a sufficient condition to achieve this. Without specific management attention, information systems are unlikely to assist in knowledge-based decision making, so managers should establish firm processes to collect and analyse data, and utilise this in an analytical manner. Managers should also make determined efforts to enhance production competence. Again, this does not happen on its own, and proactive initiatives are required to enhance an understanding among operators and maintenance staff of how technology operates.

REFERENCES

Aharoni, Y. (1993). In search for the unique: can firm-specific advantages be evaluated? *Journal of Management Studies*, *31*(1), 31–49. doi:10.1111/j.1467-6486.1993.tb00294.x

Bohn, R. E. (1994). Measuring and managing technological knowledge. *Sloan Management Review*, *36*(1), 61–73.

Coates, T. T., & McDermott, C. M. (2002). An exploratory analysis of new competencies: a resource based view perspective. *Journal of Operations Management*, *20*, 435–450. doi:10.1016/S0272-6963(02)00023-2

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning and innovation. *Management Science*, *5*(1), 227–251.

Cristofolini, I., Filippi, S., & Bandera, C. (2009). The role of product feature relations in a knowledge based methodology to manage design modifications for product measurability. *International Journal of Production Research*, *47*(9), 2373–2389. doi:10.1080/00207540701666246

Crossan, M. M., & Berdrow, I. (2003). Organizational learning and strategic renewal. *Strategic Management Journal*, *24*, 1087–1105. doi:10.1002/smj.342

Dahane, M., Clementz, C., & Rezg, N. (2008). Analysis of joint maintenance and production policies under a subcontracting constraint. *International Journal of Production Research*, *46*(19), 5393–5416. doi:10.1080/00207540802273686

Dekkers, R. (2009). Distributed Manufacturing as co-evolutionary system. *International Journal of Production Research*, *47*(8), 2031–2054. doi:10.1080/00207540802350740

Dimnik, T. P., & Johnston, D. A. (1993). Manufacturing managers and the adoption of advanced manufacturing technologies. *Omega*, *21*(2), 155–162. doi:10.1016/0305-0483(93)90048-P

Ferdows, K., & De Meyer, A. (1990). Lasting improvements in manufacturing performance: in search of a new theory. *Journal of Operations Management*, *9*, 168–184. doi:10.1016/0272-6963(90)90094-T

Galbreath, J. (2005). Which resources matter the most to firm success? An exploratory study of resource-based theory. *Technovation*, *25*(9), 979–987. doi:10.1016/j.technovation.2004.02.008

Gourlay, S. (2006). Conceptualizing Knowledge Creation: A Critique of Nonaka's Theory. *Journal of Management Studies*, *43*(7), 1415–1436. doi:10.1111/j.1467-6486.2006.00637.x

Hult, G. T., Ketchen, D. J., Cavusgil, S. T., & Calantone, R. J. (2006). Knowledge as a strategic resource in supply chains. *Journal of Operations Management*, *24*, 458–475. doi:10.1016/j.jom.2005.11.009

Irwin, H., More, E., & McGrath, M. (1998). Relationship management for innovation: the central role of communication in Australia's participation in two high-tech industries. *Technology Analysis and Strategic Management*, *10*(4), 467–481. doi:10.1080/09537329808524328

Jones, O., & Craven, M. (2001). Expanding capabilities in a mature manufacturing firm: absorptive capacity and the TCS. *International Small Business Journal*, *19*(3), 39–55. doi:10.1177/0266242601193003

Katz, R., Rebentisch, E. S., & Allen, T. J. (1996). A study of technology transfer in a multinational cooperative joint venture. *IEEE Transactions on Engineering Management*, *43*(1), 97–105. doi:10.1109/17.491273

Leonard-Barton, D. (1995). *Wellsprings of knowledge*. Boston: Harvard University Press.

Lynskey, M. J. (1999). The transfer of resources and competencies for developing technological capabilities - the case of Fujitsu-ICL. *Technology Analysis and Strategic Management*, *11*(3), 317–336. doi:10.1080/095373299107384

Macher, J. T. (2006). Technological Development and the Boundaries of the Firm: A Knowledge-Based Examination in Semiconductor Manufacturing. *Management Science*, *52*(6), 826–843. doi:10.1287/mnsc.1060.0511

Marmier, F., F., Varnier, C. & Zerhouni, N. (2009). Proactive, dynamic and multi-criteria scheduling of maintenance activities. *International Journal of Production Research*, *47*(8), 2185–2201. doi:10.1080/00207540802311106

McKone, K. E., Schroeder, R. G., & Cua, K. O. (1999). Total productive maintenance: a contextual view. *Journal of Operations Management*, *17*, 123–144. doi:10.1016/S0272-6963(98)00039-4

Meredith, J. R. (1987). Automating the factory. *International Journal of Production Research*, *25*(10), 1493–1510.

Mowery, D. C., & Oxley, J. E. (1995). Inward technology transfer and competitiveness: the role of national innovation systems. *Cambridge Journal of Economics*, *19*, 67–93.

Pintelon, L. M., & van Wassenhove, L. N. (1990). A maintenance management tool. *Omega*, *18*, 59–70. doi:10.1016/0305-0483(90)90018-5

Reagans, R., Argote, L., & Brooks, D. (2005). Individual experience and experience working together: predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, *51*(6), 869–881. doi:10.1287/mnsc.1050.0366

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: boundary spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, *22*, 287–306. doi:10.1002/smj.160

Rouse, M. J., & Daellenbach, U. S. (2002). More thinking on research methods for the resource-based perspective. *Strategic Management Journal*, *23*, 963–967. doi:10.1002/smj.256

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017. doi:10.1016/j.respol.2006.05.002

Tu, Q., Vonderembse, M. A., Ragu-Nathan, T. S., & Sharkey, T. W. (2006). Absorptive capacity: enhancing the assimilation of time-based manufacturing practices. *Journal of Operations Management*, *24*, 692–710. doi:10.1016/j.jom.2005.05.004

Uzzi, B., & Lancaster, R. (2003). Relational embeddedness and learning: The case of bank loan managers and their clients. *Management Science*, *49*, 383–399. doi:10.1287/mnsc.49.4.383.14427

Zahra, S. A., & George, G. (2002). Absorptive capacity: a review, reconceptualization, and extension. *Academy of Management Review*, *2*(2), 185–203. doi:10.2307/4134351

KEY TERMS AND DEFINITIONS

Absorptive Capacity: is a set of organizational routines and processes by which firms acquire, assimilate, transform and exploit knowledge to produce a dynamic organizational capability (Zahra and George, 2002).

Production and Process Technology: is embodied in the equipment or the means to produce a specific good, or provide a service. It is manifested in the finished product/service.

Technology Transfer: is the process by which science and technology are diffused throughout human activity. Wherever systematic rational knowledge developed by one group or institution is embodied in a way of doing things by other institutions or groups, we have TT. This can either be transfer from more basic scientific knowledge into technology or adaptation of an existing technology to a new use. TT differs from ordinary scientific information transfer in that to be really transferred, it must be embodied in an actual operation of some kind.

Technology: may be construed as differentiated knowledge about specific applications: typically the assets that form the basis of a firm's competitiveness.

This work was previously published in Knowledge Management in Emerging Economies: Social, Organizational and Cultural Implementation, edited by Minwir Al-Shammari, pp. 66-78, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 12
Tracing the Implementation of Non-Functional Requirements

Stephan Bode
Ilmenau University of Technology, Germany

Matthias Riebisch
Ilmenau University of Technology, Germany

ABSTRACT

A software architecture has to enable the non-functional properties, such as flexibility, scalability, or security, because they constitute the decisive factors for its design. Unfortunately, the methodical support for the implementation of non-functional requirements into software architectures is still weak; solutions are not generally established. Recently, there are only few approaches that actually deal with non-functional requirements during design; even fewer take advantage of traceability, which supports a mapping of requirements to solutions through the development process. Therefore, in this chapter the new architectural design method TraGoSoMa is presented, which supports these issues. The method uses a so-called Goal Solution Scheme, which guides the design activities, supports conflict resolution, decision-making, and the classification of solutions. For illustration purposes the chapter uses a case study from a reengineering project for a Manufacturing Execution System (MES) that is restructured according to the SOA principles and integrated with an Enterprise Resource Planning (ERP) system.

INTRODUCTION

Motivation

Performance, scalability, flexibility, security and other so-called non-functional requirements or quality properties are crucial for the success of nearly every software project. They bear even more risk than functional requirements, because they can hardly be implemented after making the major design decisions. The software architecture has to enable these non-functional requirements, because they constitute the decisive factors for its design. Several architectural methods emphasize the analysis of non-functional requirements (Hofmeister et al., 2000) and the design considering them (Bosch, 2000; Bass et al., 2002). Furthermore, in the field of requirements engineering

DOI: 10.4018/978-1-4666-1945-6.ch012

functional and non-functional requirements and their interdependencies are modeled using some mature and established goal-oriented approaches (Chung et al., 2000; Yu, 1995; Amyot, 2003). Therefore, some development steps of requirements engineering and architectural design are strongly related with each other. However, bridging the gap between these two research areas is still a critical issue (Galster et al., 2006) especially in the case when non-functional requirements change.

Service-oriented architectures are frequently applied for business-critical purposes. For these systems a long life expectancy constitutes an important concern, during which they have to be adjusted to a high number of changes to maintain their operation and their business value. Thus, evolvability is an important issue for this type of systems.

Traceability links support changes, and therefore the evolution of software systems, by expressing relations between artifacts in different phases of software development (Letelier, 2002). They facilitate program comprehension by expressing dependencies explicitly. They relate design decisions to constraints, and they are used to trace dependencies for checking the completeness of changes. These benefits can be achieved from fine-grained traceability links on the level of design elements and design decisions.

There is a considerable amount of research for managing traceability and for maintaining their accuracy during changes, which is for example discussed in (Mäder et al., 2006). However, the tracing of non-functional requirements usually requires a very high number of links, since typically a high number of a system's components depend on one such requirement.

There are some critical issues that make the whole process complex: for example (a) the mapping of high-level non-functional requirements to their low-level solutions and mechanisms, as well as (b) detecting and solving conflicts among non-functional requirements resulting in trade-offs, and further (c) the missing structuredness of existing methodologies or even their complete absence.

The management of traceability links, and with them the non-functional properties, requires a high human effort, first because of the high number of links and second because of missing rules inhibiting effective tool support. In order to reduce the effort for establishing, maintaining, and validating the traceability links, our work presents contributions to facilitate tool support, and hence, traceability of non-functional requirements, in several ways, which are described in the next sections.

Challenges

The proper treatment of non-functional requirements is a very important part of architectural design. Unfortunately, the methodical support for the implementation of non-functional requirements is still lacking even if this book addresses it. The absence of a consolidated set of solutions for non-functional requirements, the abstraction gap between requirements and design as well as the many influencing factors on design decisions reduce the applicability of detailed design instructions.

From the point of view of software architectural design there are only few methods that lead to an improved design process by especially considering non-functional requirements: for example the QASAR method (Bosch, 2000) or the Attribute-Driven Design (ADD) method (Bass et al., 2002). On the other hand, in requirements engineering there are adequate approaches for dealing with non-functional requirements in a goal-oriented way: for example the NFR framework (Chung et al. 2000), the i* framework (Yu, 1995), or the Goal-oriented Requirement Language (GRL) (Amyot, 2003). Although efforts are made to push these goal-oriented methods towards support for architectural design, and therefore some overlapping in the development steps can be observed, further research has to be done to bridge the gap.

The goal-oriented techniques are valuable for dealing with the interdependencies and the refinement of functional and non-functional requirements, and to some extent are able to map high-level solutions to those goals. However, despite some research on their support for architectural design, (e.g., Grau & Franch, 2007 or Liu & Yu, 2001), on their own they are rather unsuitable and insufficient for the whole architectural design process. Therefore, from our point of view, we have to adjust and integrate these techniques into the software architectural methodologies and publish this knowledge to software architects. This will help them as a means for going from the problem space to the solution space and for choosing and evaluating proper architectural solutions especially for non-functional requirements.

As another aspect, traceability promises strong benefits for the design and evolution of service-oriented architectures and applications especially if traceability links are available on a fine-grained level. However, this leads to a very high number of traceability links. Thus, the complexity of the link structure and a high overhead effort for its maintenance constitute hampering factors. This holds especially for the traceability of non-functional properties, since usually a high number of design elements depend on each non-functional requirement, and this results in a high number of traceability links related to these requirements.

The integration and application of the above-mentioned methods provides the potential for an improved traceability. However, the traceability concepts have to be integrated within these methods, and vice versa the methods have to be adapted to make use of traceability. As a next step after integration, a proper tool support for the management of the traceability links has to be built, covering the establishment, the maintenance, the validation, and the exploitation of the links.

Objectives and Contribution

In this work we present the design method TraGoSoMa (an acronym for Traceability-driven Goal Solution Mapping) for a proper treatment of non-functional requirements during architectural design. The method is intended to give a better support and guidance for architectural design activities while dealing with non-functional goals. It supports the conflict resolution between competing goals and a classification of solutions to the goals they are derived from, as well as a systematic selection based on methodical decision-making. TraGoSoMa clearly is a software architectural design method, even if it partly integrates requirements engineering activities and goal-oriented techniques for modeling non-functional requirements. With its Goal Solution Scheme as a core concept the method maps non-functional goals and subgoals to architectural principles, which are essential criteria for finding the right architectural design, and it maps them further to the functional and technical solutions. Therefore, it enables an accumulation of architectural know-how. Beyond, the application of the method facilitates traceability link establishment, and thus, a better maintainability of the software. The method is illustrated in the next sections with the help of a case study.

BACKGROUND

In this section we briefly investigate the contributions of state-of-the-art methods for non-functional design. Therefore, we discuss Bosch's Quality Attribute-based Software Architectural (QASAR) method (Bosch, 2000) and the Attribute-Driven Design (ADD) method (Bass et al. 2002). We express how our approach is influenced by the goal-oriented approaches from requirements engineering, such as the Non-Functional Requirements (NFR) framework of Chung et al. (2000) and the i* framework which was introduced by

Yu (1995). Moreover, we relate our traceability framework for non-functional requirements to other works concerning traceability.

The Design Method QASAR

A frequently performed way of constructing a software architecture for both kinds of requirements is described by Bosch's QASAR method (Bosch, 2000). QASAR considers non-functional requirements by architectural transformations. The method describes three phases. In the first phase, the functional requirements are implemented by functional components with the Functionality-based Architectural Design (FAD) method. FAD uses core abstractions of functional concepts—the so-called archetypes—to derive architectural components by functional decomposition. In the second phase, the developed architecture is assessed in order to decide whether the non-functional requirements are fulfilled or not. Different approaches for the assessment of the non-functional requirements can be used in this phase, e.g., scenario-based evaluation, simulation, mathematical modeling, or objective reasoning. Once the non-functional properties of the architecture are assessed, in the third phase the architecture is transformed to satisfy the non-functional requirements specifications. This transformation leads to suitable functional structures and components, which are developed for the implementation of as many as possible non-functional requirements. All remaining non-functional requirements are implemented by changing all affected components. In this phase the changes are scattered over the system.

We consider QASAR as a very important method of architectural design and we will use its core concept—the fulfilling of non-functional requirements by functional solutions—in our TraGoSoMa method. The drawback of QASAR, the scattered implementation of the remaining non-functional requirements, results from its third phase. The scattering is accompanied with a demand for a very high number of traceability links and hampers maintainability. We want to address this issue in the TraGoSoMa method by a careful consideration of the non-functional properties.

Attribute-Driven Design

The ADD method of Bass et al. (2002) is a step-by-step method for designing software architectures. ADD considers non-functional requirements at least as important as functional ones and takes them as input together with other constraints. The non-functional requirements, or quality attributes, have to be specified as quality attribute scenarios before. Then, with ADD the requirements are transformed into a conceptual architecture by a recursive refinement, which is applied in several design steps.

First, an architectural element is chosen to be refined, beginning with the system itself. Next, those requirements that influence the architecture the most have to be determined. These architecturally significant requirements—called architectural drivers—are quality, business, or functional goals and can be identified by prioritization. In a further step, architectural styles and patterns are chosen that satisfy the architectural drivers. In this third step, for the one architectural element that is to be refined and its high-prioritized architectural drivers, the most appropriate solution has to be chosen. The fourth step is about the actual refinement. The chosen style is instantiated and corresponding architectural elements are created. The new elements are assigned with functionality they are responsible for, and their interfaces are specified for information flow. In the fifth step, the refinement of the child elements is prepared by verifying and refining their requirements descriptions. Finally, all steps are repeated for further decomposition until all architectural drivers are fulfilled.

ADD, just as QASAR, achieves the fulfilling of quality goals by allocating functional components. Its strength is that it concentrates on the most significant architectural drivers for choosing a proper functional solution utilizing appropriate architectural styles and patterns. We will use this concept in our TraGoSoMa method by prioritizing requirements. However, ADD also has its drawbacks. First, the method ends with only a conceptual architecture and no concrete components; hence, the design process has to be continued by other means. Furthermore, ADD does not exactly describe how to identify the architectural drivers. This step is apparently left for the requirements engineer. Additionally, due to its recursive nature, the capability of the ADD method to evaluate alternative solutions for the overall architecture is limited. If an architectural element is refined once and the method continues with the child elements, there is no possibility to consider an alternative for the previous decision. Besides, there is nothing about traceability support with ADD, which is an important issue we want to address with our TraGoSoMa method.

Goal-Oriented Approaches

In the field of requirements engineering goal-oriented approaches are a popular way to elicit and model requirements. The Non-Functional Requirements (NFR) framework by Chung et al. (2000) constitutes one of these approaches for dealing with non-functional requirements. The framework uses the concept of so-called softgoals that represent non-functional requirements. Softgoals are goals that have no clear-cut definition or criteria for their satisfaction. Development decisions often contribute only partially to or even against these goals within acceptable limits. Because softgoals are accomplished rather partially than completely or not at all, they are considered as fulfilled, and then called *satisficed*, if an adequate level of the criteria is reached. The interdependencies between softgoals can be categorized into several contribution types according to their weak or strong positive or negative influence on satisfying a softgoal.

The NFR framework describes several interleaving and iterative activities to arrange softgoals with their interdependencies in a Softgoal Interdependency Graph. First NFR softgoals are established and refined into subgoals by decomposition. Then, different architectural solutions are developed as so-called operationalizations for the softgoals. Furthermore, softgoals can be refined by so-called argumentations, which model domain characteristics and developers' expertise. Beyond, during the refinement, decisions about the criticality of different goals can be made. Performing these activities, different alternative solutions and corresponding design trade-offs are elaborated in a well-founded way. Implicit interdependencies among softgoals are detected and documented together with design rationale. Alternative solutions are evaluated regarding their contribution to the non-functional requirements. Finally, an adequate solution is selected based on the evaluation.

Another important goal-oriented method is the i* framework, which was presented by Yu (1995). Akin to the NFR framework, i* uses the notation of goals and softgoals to model system requirements. In addition, it contains further elements, e.g., tasks, resources, and actors, and it uses different types of links, as for example contribution, decomposition, and means-end. Therefore, it can be utilized to not only describe the system's requirements but also the organizational context.

In 2008 the User Requirements Notation (URN) (Amyot, 2003) was approved as an international standard as ITU-T Recommendation Z.151. URN consists of the Goal-oriented Requirement Language (GRL), which is based on NFR and i*, and Use Case Maps (UCM), which are used for scenario modeling. GRL as an elaborate notation can be used especially to model non-functional requirements as goals.

We consider the described goal-oriented approaches as appropriate means for modeling non-functional requirements. By explicitly considering non-functional requirements right from the beginning and providing a well-defined evaluation process, they can help to design software architectures. Therefore, our TraGoSoMa method builds on them and uses their notation of softgoals for non-functional requirements and contribution links for the interdependencies.

However, the approaches focus on the refinement of non-functional requirements and their analysis from a requirements engineering point of view. From an architectural point of view, despite some works also dealing with architectural design as for example the one of Grau and Franch (2007), further design activities and notations have to be used. We argue that there are several drawbacks. For example, in the goal-oriented approaches, architectural constraints from the environment as well as interdependencies with technical solutions are not considered. They cannot assure that a selected solution can be implemented with a specific technology. Furthermore, the goal-oriented approaches do not consider general design principles in their decomposition and refinement, although they are very important for choosing proper solutions during architectural design. Beyond, the comprehension of the quality goals' background is an important issue for architectural design. That is way giving only goal models as input to software architects as a means for description is rather insufficient. Therefore, we integrate these techniques in a context of further design steps. Furthermore, in the work of Chung et al. (2000) they only concentrate on accuracy, performance, and security requirements—on the last one in a rather outdated way—and are not concerned with traceability issues. We will later discuss traceability and also some security aspects, when describing our TraGoSoMa method, in more details.

Traceability

Traceability is a concept to relate artifacts from different development stages, such as requirements analysis, design, and implementation via links. The use of traceability links is advantageous even if additional effort is required. Traceability links facilitate system comprehension by providing information about dependencies between artifacts and entities. Requirements traceability enables to trace back the origin of a requirement to its elicitation and to document every change made to it. Important works in this field are the ones of Gotel and Finkelstein (1994), Pohl (1996), Ramesh and Jarke (2001), Letelier (2002) as well as Pinheiro (2004). Other approaches for traceability link establishment consider links between requirements and test cases, for example the scenario-driven approach by Egyed (2001) and the approach by Olsson and Grundy (2002).

In the scope of our work, traceability links can be used to trace design decisions during the development process. Approaches regarding traceability from requirements to design and between various design artifacts have to be considered. Traceability links for design shall be established and adapted while building or manipulating models or other development activities. Therefore, the steps of link establishment have to be embedded into the steps of design methods. We have to state that—to the best of our knowledge—there are no approaches of this kind. For establishing traceability links regarding non-functional requirements and design artifacts there is the probability-based retrieval approach by Clelang-Huang et al. (2005) called Goal-Centric Traceability. Their user evaluation step to discard incorrectly retrieved links to increase precision is valuable. However, they can only identify links that are incidentally included in the descriptions of artifacts that were elaborated regardless of traceability. They cannot identify links, if the investigated artifacts do not cover the relations. Therefore, the completeness and correctness of the links can never be optimal. The incremental approach of Latent Semantic Indexing by Jiang et al. (2007) aims at the identification of related elements with link recovery. It can be helpful for finding links in existing designs and maintaining their change, but it cannot provide all links.

For the definition of relations between design activities and traceability links, a standardized definition of the syntax and semantics of the traceability links is necessary. Unfortunately, the definition of a standard set of traceability link types is still an unresolved issue. Due to different research goals, a high number of traceability link types has been defined, for example in (Pohl, 1996) or (Ramesh & Jarke, 2001). As a step towards simplification and abstraction, we will later restrict ourselves on a small set of types.

For an exploitation of the traceability links, they have to be established on a detailed level, for example to generate test cases; they have to connect model elements and expressions. As a result, the number of links and the complexity of traceability information are high. The issue of maintaining and checking the links leads to a high effort and thus tool support is essential. An overview of research topics, results, and open issues in the field of traceability is discussed in (Mäder et al. 2006) and (Winkler & von Pilgrim, 2010).

There is already support for traceability by requirements management tools, e.g., Requisite Pro or Doors, and by goal-modeling tools, such as jUCMNav (Roy et al. 2006). However, their support for linking other artifacts than requirements is limited. Mäder et al. (2008) present an approach that tackles the problem of automated traceability for UML-based development. Their *traceMaintainer* is a rule-based prototype tool for traceability link maintenance. Nonetheless, there is a need for further work, because the tool does not cover initial link establishment and tool support should not be limited to UML models as well. For an automated link establishment, the definition of proper rules is an important issue. Therefore, as a first step, in (Bode & Riebisch, 2009) we already described in detail, how traceability links can be established during several development steps and between different design artifacts using the example of the category-based design methodology Quasar (Siedersleben, 2004).

In the next section we illustrate our design method TraGoSoMa for tracing the implementation of non-functional requirements. We establish traceability links between non-functional requirements, architectural principles, and functional solutions. Therefore, we present a set of link types and their semantics, and we explain how they are applied while performing the several design activities according to our Goal Solution Scheme. With the help of these traceability links according to the Goal Solution Scheme, especially the comprehension of the transformation of non-functional requirements to solutions is supported. The design-decisions are traced and rationale for the decisions can be connected to the links. Furthermore, with the help of the links and an evaluation of the decisions, the completeness of solutions for the non-functional requirements can be checked.

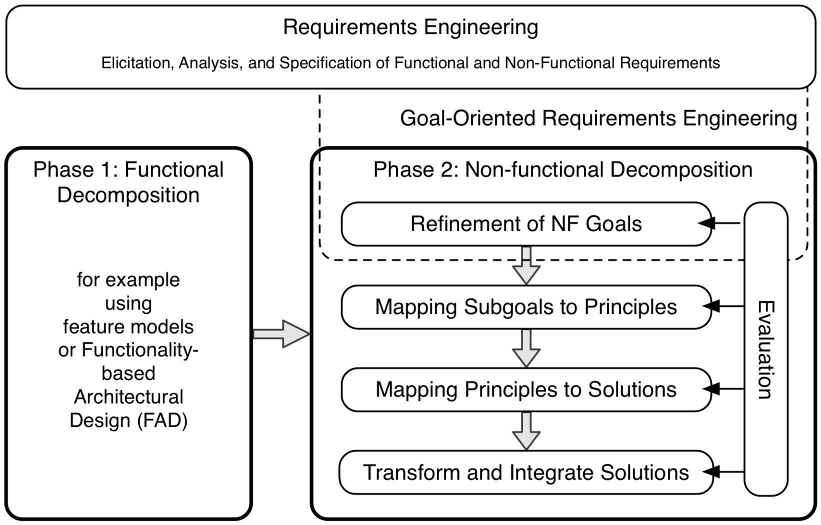
ARCHITECTURAL DESIGN PROCESS OF THE TRAGOSOMA METHOD

In this section we present the design method TraGoSoMa, named by an acronym for Traceability-driven Goal Solution Mapping. We first give a short overview of the method; afterwards we introduce our case study, which illustrates the design activities, and the Goal Solution Scheme, which constitutes a core concept and drives our method. Further, traceability issues are discussed and the phases of the design method are explained in detail.

TraGoSoMa Method Overview

The method TraGoSoMa represents an architectural design method with focus on non-functional properties. It starts after elicitation, analysis, and specification of the requirements and consists of different phases and steps illustrated in Figure 1. First of all, a functional decomposition of the sys-

Figure 1. Phases and steps of the TraGoSoMa method related to requirements engineering



tem is performed akin to QASAR's FAD method. This could be done using different approaches, such as feature models or function trees, and results in candidates for architectural components, the responsibilities of which are determined by functional requirements. We propose to perform decomposition according to the functional requirements first, because this is for example necessary to determine the security-relevant parts of the system; see (Bode et al. 2009) for details. However, designing for the non-functional goals is even more important, but less understood today. So, in the second phase of our method especially the design for non-functional requirements is handled.

This second phase is strongly connected to the Goal Solution Scheme. The first step is performed, if not already provided from the requirements specification. The soft, imprecise, mostly vague, ambiguous, and competing non-functional requirements, such as flexibility, scalability, or security, are modeled as goals and refined into subgoals. This enables the resolution of conflicts between the goals by prioritizing the subgoals accordingly. In this step, we utilize the goal-oriented approaches discussed before, instead of pure scenario-based techniques as proposed by most architectural design methods. In a second step, these subgoals are mapped to the design principles, which support the subgoals. We introduce this mapping, which is not performed in the goal-oriented approaches, because general design principles are very important for choosing proper solutions during architectural design. Additionally, the architect has to be involved in this modeling process since only providing goal models as input for further design is not enough. In the third step, possible functional solutions and technical components, as so-called solution instruments that support specific principles and subgoals are identified, and traceability links are established accordingly. In this second and third step several architectural decisions have to be made about which principles and solutions are considered for design of the software architecture with regard to trade-offs and synergies between different subgoals. These decisions are significant, because the effect of the relations between subgoals can be mutual enhancement or reduction. A final fourth step in the second phase further combines the functional decomposition of the first phase with the findings from the second.

As a result of the second phase, the architectural components are defined according to functional and non-functional responsibilities, and functional solutions are elaborated for non-functional responsibilities. The solutions and technical components established in the second phase with the help of the Goal Solution Scheme are integrated into the software architecture. Therefore, adequate architectural transformations have to be performed akin to the third phase of QASAR. The kind of these transformations, and how they are applied, is discussed in a later section. All activities in the second phase of the TraGoSoMa method have to be subject to an early assessment within an iterative design process. With the evaluation all decisions and the contributions of the solutions to the non-functional requirements are checked. We deal with this issue in an extra section.

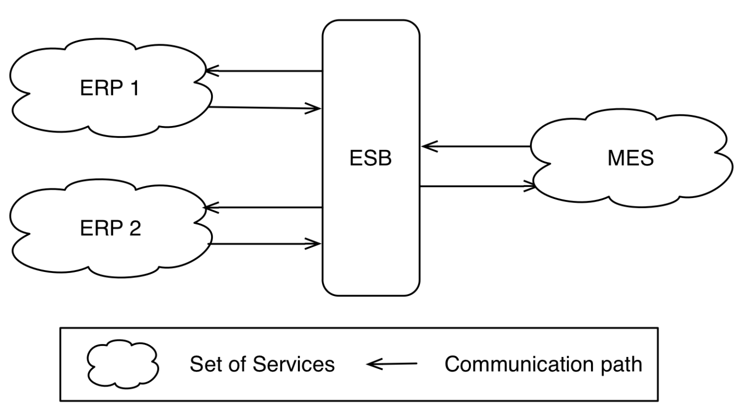
In the following sections we assume the requirements specification and functional decomposition to be done and concentrate on the second phase of the architectural design process dealing with the non-functional properties. We also explain how this approach facilitates design traceability.

Service-Oriented MES Case Study

For the illustration of the TraGoSoMa method we use a case study from a reengineering project for a Manufacturing Execution System (MES) that is restructured according to the SOA principles. A MES manages the manufacturing in modern flexible plants (VDI, 2007). It is connected to Enterprise Resource Planning (ERP) systems, which handle manufacturing plans and actions and represent a business perspective, while the MES is able to manage the manufacturing actions on a more fine-grained level. An MES has access to the abilities and the limitations of the real manufacturing processes and, therefore, it is able to optimize them, and simultaneously provides an increased flexibility. The MES covers tasks, such as detailed scheduling and process control, the management of machines, material, and personnel, etc. For the case study we focus on the integration between ERP and MES. The requirements to the interface between both are defined by the ANSI standard ISA-95 (ISA, 2000). As the platform for the MES interface the Enterprise Service Bus (ESB) (Chappell, 2004) has been chosen, in a style similar to a middleware. Figure 2 shows the integration interface and its environment.

There are some non-functional properties an MES has to fulfill: a high flexibility and scalability, time behavior as well as security. As an example for a security requirement we mention the information flow control. In our case it is important to protect the business-critical private information of the customer 1 from an unauthorized access by its competitor customer 2. Even if

Figure 2. Overview of the integration interface



both give manufacturing tasks to the producer, no details about the order must be disclosed to the competitor through the ESB or the MES, for example details concerning amount, specification, and technological process. Flexibility is necessary regarding different planning algorithms, control principles, and regarding the integration with a variety of machines and ERP systems. The requirements for scalability arise from the need for mastering complex manufacturing tasks with a high number of variants and elements, and for the interoperation with multiple different ERP systems due to outsourcing.

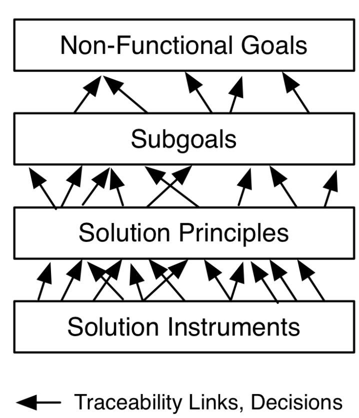
Goal Solution Scheme

The Goal Solution Scheme was developed to guide the designer to deal with non-functional properties while designing the architecture, to ease resolving conflicts between competing goals and design principles, and to facilitate decision-making. It has some similarities with goal-graphs from requirements engineering. However, the scheme extends it by the explicit consideration of design principles and a classification of functional and technical solutions. As indicated in Figure 3, the Scheme maps non-functional requirements to their subgoals, to solution principles supporting these goals, and further to solution instruments, such as technical solutions and components, for an implementation of the principles. The mapping is represented by traceability links, which are visualized in Figure 3 by the arrows. From top to bottom the scheme represents possible refinements and decisions during the implementation of non-functional requirements. The traceability links carry the information about the design decisions. By providing guidance the scheme refines the design methods and facilitates the establishment and maintenance of traceability links related to the design activities.

The scheme guides the activities within analysis and design and represents them by transitions between the layers. The upper transition supports the resolution of conflicting non-functional properties by a refinement to subgoals. Such a specification and refinement of non-functional goals can also be provided by the goal-oriented approaches we discussed above, e.g., the NFR framework, i*, or GRL. However, we included this transition step because this kind of specification of non-functional goals is seldom provided to software architects in practice. The comprehension of this transition is necessary for the software architects as well, because they have to contribute to the prioritization of the non-functional goals and they have to implement the goals by selecting proper principles. Since we do not want to reinvent the wheel, we adopt the goal-oriented notation for softgoals and the contribution links, even if this leads to an overlapping with requirements engineering models.

The next transition of the scheme guides the designer from non-functional goals to solution principles, which is a novelty step regarding the design progress towards a solution in comparison to the goal-oriented requirements engineering approaches. In the lower transition of the scheme, technical solutions and existing components are mapped as solution instruments to the principles, making decisions on how to implement the goals and principles by solutions. A similar concept is used for the Factor-Strategy Models of Marinescu

Figure 3. Structure of the Goal Solution Scheme



and Ratiu (2004) who use design rules and principles to map metrics to quality goals.

As a major contribution the scheme facilitates the prioritizing for decision-making. Furthermore, the scheme supports the resolution of conflicting non-functional requirements by an identification of potential trade-offs and synergies. It provides a fine-grained sequence of design steps. In this way, the scheme represents a refinement of the design activities, which are represented by traceability links. A reduction of the traceability link complexity is achieved, since one or a few principles and solution instruments implement one subgoal. In an ad-hoc design non-functional requirements are implemented in a scattered manner leading to a much higher number of links, for example 10 or 100 times higher.

The Goal Solution Scheme constitutes the central concept for the simplification of the traceability links. It facilitates the traceability in several ways:

- Significant reduction of the number of links,
- Guidance for the designer by sequences of proposed design activities, which enables tool-supported decisions and automation,
- Simplification of link checks for accuracy, completeness, and consistency by a comparison between chosen solutions and relations within the Goal Solution Scheme.

As a result, the scheme provides an alignment of solution principles and solution instruments, it classifies them according to their impact on non-functional properties, and it provides a stock of reusable solution instruments to the architect and the designer. The solution instruments serve as a source of proposals for design alternatives during decision-making.

The scheme is not a design artifact that requires additional maintenance and effort. It rather is a data structure on a meta-level, which indicates and guides how the architectural design steps should be performed in the TraGoSoMa method. If traceability links are established appropriately and managed in a repository, the organization of the repository reflects the Goal Solution Scheme. Beyond, a software architect can use the scheme as a knowledge base, where solution instruments are mapped to principles and goals.

Traceability Links

As mentioned previously, traceability links connect artifacts in the sequence the developer has built or accessed them. Furthermore, they carry the information about the design decisions that lead to the related solution instruments, such as components. For the link tracking, evaluation, and exploitation, the type of the traceability link is important, because it determines a link-semantics as mentioned in the background section; the number of link types should be small.

Additional information can be stored attached to the link, e.g., design decisions. Important elements an explicit traceability link comprises are:

- a unique identifier for its recognition and to avoid ambiguity,
- a start element as source of the link, including type and context of this element,
- an end element as destination of the link, including type and context,
- the type of the link.

For the decision-making, traceability links are advantageous because they make decisions explicit and comprehensible. Alternatives can be estimated. Software architecture design decisions should always be documented with their design rationale (Clements et al. 2003). Therefore, they are recently seen as first class entities. Duenas and Capilla (2005) for example introduced a decision view for software architectures. The Goal Solution Scheme facilitates the tracing of such decision entities to their related artifacts. Therefore, a link may contain additional information:

- a reference to a design rule for this specific activity,
- the decision connected with the development activity, including the goal of the decision, alternatives, the rating of the alternatives and the choice,
- the link status concerning the certainty of correctness (e.g., after changes of the connected elements or during reverse engineering activities),
- the creator of the link,
- a temporary priority to control the tracking of the links.

As introduced by earlier works (Mäder et al., 2007), we distinguish four different link types, which are sufficient for the most design situations:

- *refine* – for an activity increasing the level of detail, either by specialization or by decomposition including the AND and OR types.
- *realize* – represents a step towards the solution (e.g., between a non-functional goal and a design principle or between a principle and a solution)
- *verify* – compares the behavior and the properties of requirements and of the developed solution or its parts (e.g., between a use case and a test case) and
- *define* – relates the establishment of an identifier and its usage.

Additional link types can be introduced for dependency types from utilized models as for example UML models. In order to be able to handle the goal models and their contribution links we added the link type

- *contribution* – which can express degrees of positive or negative influence between model elements.

With the help of the contribution links an evaluation of alternatives can be achieved, for example in relation to the goal models. However, positive contributions can also be linked as realizations, when a solution is chosen. The realization links then enable tracing which path was taken from the problem space to the solution space and how the goals are achieved by the implemented solution.

Traceability links can be tracked in both directions, regardless of the direction that is defined by the link type. Besides, a distinction between implicit and explicit links is necessary. Explicit traceability links are established, while a developer performs a software development activity. Implicit traceability represents existing associations between elements of the system model using identifiers, for example between an analysis and a design artifact. These traceability links are references, but they are evaluated if traceability links are tracked during their utilization.

The type of the traceability links is defined according to the TraGoSoMa design activities. The transitions in the Goal Solution Scheme represent these activities. In the first transition the non-functional goals are decomposed and thus linked to the subgoals by links of the type *refine*. In the second and third transition the positive and negative influence has to be represented. Therefore, links of the type *contribution* are established. In addition to the influences, these transitions represent steps towards solutions. Consequently, the link type *realize* is used to express this aspect. A possible decomposition on each level of the Goal Solution Scheme, for example a decomposition of solutions, can also be traced by links of the type *refine*.

Goal Refinement and Elaboration

The several non-functional requirements for a product are often competing and conflicting in their interdependencies. As a solution they have to be prioritized. If this is not possible on the top-level, a resolution is attempted after a refinement. By refining the requirements vague interdependencies can be concretized and previously hidden dependencies can be made explicit. To determine the mutual impact of the relations and to detect conflicts and synergies, the requirements or goals can be classified in dominating (fundamental) and supporting (instrumental) ones (Wohlfarth, 2008).

The refinement of the non-functional goals is covered by the first transition in the Goal Solution Scheme. This is the first step in the second phase of our design approach. The refinement of the non-functional properties is necessary for their comprehension and makes them more specific. It can be performed according to the abovementioned goal-oriented approaches, e.g., the NFR framework.

For the refinement standards can help, for example the ISO 9126 (ISO, 2001). This standard for instance provides subgoals for maintainability, namely analyzability, changeability, stability, testability, and maintainability compliance. Furthermore, the Goal Question Metrics (GQM) method can be applied to identify subgoals. This structured querying technique helps to analyze influences on a goal (Basili et al., 1994). The Factor-Strategy Model of Marinescu and Ratiu (2004) uses a similar principle for mapping quality goals to metrics.

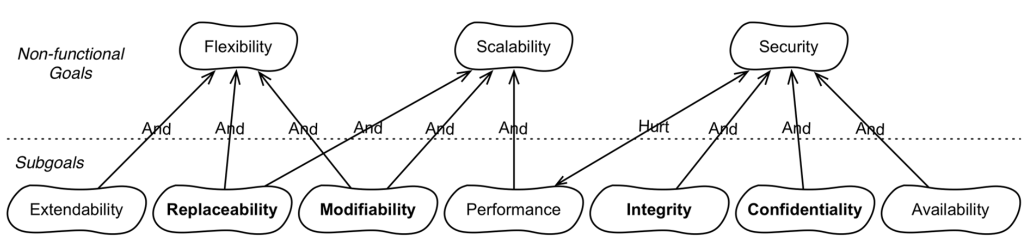
According to the NFR framework (see section *Background*) non-functional requirements have a type and a topic. As an example, the requirement "Security of accounts" has the type "security", which indicates the specific NFR, and the topic "accounts", which targets at the subject. Non-functional requirements can be refined regarding type or topic. The refinement of maintainability mentioned above is a refinement regarding the type of the NFR.

In our case study the top-level non-functional requirements are *flexibility*, *scalability*, and *security*. According to the ADD method they constitute architectural drivers. After a consideration of their relationships on this level we can assume that flexibility and scalability are in a rather synergetic relation to each other, because they both deal with change, while security might be conflicting to the others, because it implies restrictions of the information flow and the data access. This guess has to be proven in the next steps. For a precise analysis and a solution for our MES project, a refinement has been performed. Figure 4 shows parts of the results. The refinement is presented using the i* notation for softgoals and links.

For flexibility, there is a definition in the IEEE standard glossary of software engineering terminology (IEEE, 1990), although a detailed discussion of its subgoals is missing. Regarding flexibility some discussion can be found in the literature (Zeng & Zhao, 2002; Nelson et al., 1997; Eden & Mens, 2006; Morgan, 2006). We elaborated the subgoals *extendability* (IEEE, 1990), *replaceability* (ISO, 2001), and *modifiability* (Bengtson et al., 2004) for our example. We focus on the latter two because of their high priority.

Scalability is lacking a definition by a standard; however, some works discuss this quality attribute (Hill, 1990; Bondi, 2000; Duboc et al., 2006; Duboc et al., 2007). Scalability is always

Figure 4. First transition of the goal solution scheme (cut-out from the case study)



concerned with *performance,* or efficiency in terms of the ISO 9126 (2001), and how well a solution to a problem will work when the size of the problem increases. However, if a system performs well it is not necessarily scalable, too. Therefore, we considered replaceability and modifiability as subcharacteristics of scalability as well. If an MES has to face changes for example due to an increasing complexity of the manufacturing tasks, modifications are necessary. Moreover, it should be easy to replace parts of the whole system with more efficient ones, if this is necessary to scale up and retain a high performance.

For security there are several definitions from the International Organization for Standardization (ISO), e.g., (ISO, 2001; ISO, 2005), and Chung et al. (2000) comprehensively discuss its refinement in their NFR framework. The most important subgoals are *integrity*, *confidentiality*, and *availability*.

Before the refinement, we already mentioned our assumption for a synergetic relation between flexibility and scalability, as well as the possible conflict between scalability and security. The conflict could neither be verified nor solved on the top level, because both scalability and security are essential. However, after the refinement of the non-functional goals illustrated in Figure 4, we can try to solve the conflict and verify the synergetic interdependency between flexibility and scalability. The latter could be verified by the mutual positive influence of replaceability and modifiability on both top-level goals. Beyond, a negative interdependency between performance and security was detected, because security mechanisms, as for example encryption, require extra operations, often are time consuming, and can hamper performance. This confirms the conflict; however, for a resolution a further refinement to principles is necessary. For the further design process of our case study we will concentrate on the subgoals replaceability, modifiability, as well as integrity, and confidentiality because of their high priority.

The abovementioned refinements are expressed using *and*-contribution links according to the i* notation. In the Goal Solution Scheme they are represented by traceability links of the type refine. The *hurt*-contribution can be traced with links of the type contribution if necessary.

Decision about Solution Principles

After the first step of the second phase of TraGoSoMa, the top-level goals are refined into subgoals and are more specific. But, they are still non-functional and still cannot be implemented directly. In the second step, the transition from the subgoals to the design principles is performed, as presented by the Goal Solution Scheme.

As a step from the problem space to the solution space, in this second step, design principles and guidelines are assigned to the subgoals. These principles and guidelines give hints or advice for the functional solutions. Of course, lots of principles exist and even more relations between non-functional goals and these principles are imaginable. Therefore, the designer has to analyze the subgoals and to decide on suitable principles. It is always the case, that there are different non-functional goals that have symbiotic relations or in contrary compete with each. In order to resolve conflicts, knowing about the interdependencies between the different subgoals is important. A goal model contains these dependencies and the trade-offs.

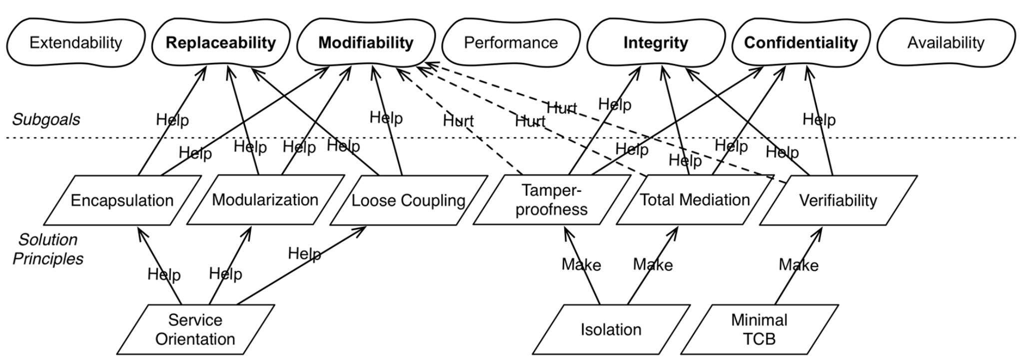
For illustration an example for a decision is discussed here. The principle of high encapsulation supports changeability. On the other hand, a strong encapsulation has a negative influence on testability, because inaccessible attributes are hard to control. Because of the refinement from the first step, both changeability and testability are known to be subgoals of maintainability and contribute to it. Now, by assigning encapsulation to these subgoals the trade-off becomes visible and can be considered. Frequently, multiple different principles contribute to the same subgoal. In these cases a decision can be made, which principle is applicable or how to prioritize them.

Trade-offs between different non-functional subgoals and solution principles often are still not tangible enough. Then, they have to be elaborated further on the solution instruments level of the Goal Solution Scheme. This is necessary to be able to decide with clear rationale, which principle to choose to achieve the highest degree of goal fulfillment. In this case, the principles are mapped further to solution instruments and the decision-making is postponed to the next step, when the criteria for adequate solution instruments are more precise than those for the principles. Based on the solution instruments' contributions to the principles and the non-functional goals, the different alternatives can be weighed and the decisions, which alternatives to choose, can be made. For the goal-oriented approaches some evaluation techniques exist, anyway, it is always reasonable to decide as soon as possible to reduce further effort.

For our case study, the second transition of the Goal Solution Scheme is partly shown in Figure 5. The subgoals result from the refinement in the previous step of TraGoSoMa. Starting from the higher prioritized subgoals, appropriate solution principles are chosen. For the subgoals replaceability and modifiability, we decide in favor of the architectural design principles *encapsulation*, *modularization* and *loose coupling*. These principles are well known to support changes. Already Parnas (1972) discussed the importance of modularization for changeability and flexibility, which is one of our most important non-functional goals. Moreover, *service orientation* was identified to support encapsulation, modularization, and loose coupling. A service-oriented architecture obviously can help in this scenario, because loose coupling is one of its core principles. It further helps encapsulation and modularization. In addition to the contribution links shown in Figure 5, the mentioned principles are explicitly related to the subgoals replaceability and modifiability by traceability links of the type realize. This type of links is established, because the principles represent a step towards the solution of the non-functional goals, and to document the design decisions for choosing service orientation.

The security subgoals integrity and confidentiality are discussed as another example. To integrate such requirements, security policies have to be applied, as a comprehensive set of rules that are designed to achieve the system's security goals (Goguen & Meseguer, 1982). Security policies are applied to determine a so-called *trusted computing base* (TCB) (Lampson et al. 1992). The TCB comprises the functional parts of a system that enforce and protect the security policy. For the implementation of a security policy and a trusted computing base, there are fundamental principles that refer to the so-called reference

Figure 5. Cut-out of the transition from subgoals to principles for the case study



monitor concept (Anderson, 1972). A reference monitor must be tamperproof, always invoked and small enough to be analyzable and verifiable, which is represented by the principles *tamper-proofness*, *total mediation*, and *verifiability*. These reference monitor principles are further supported by *isolation* and a *minimal TCB* as principles for the architectural design. Isolation of the security relevant functions in the security architecture of a system is a necessary consequence to be able to realize a tamperproof reference monitor that cannot be bypassed (Gasser, 1988). Correctness and completeness are additional necessary properties not further discussed here (Department of Defense, 1985). These decisions and the causes are again documented by traceability links of the type realize.

However, in this design step, conflicting relations between the security principles and the subgoal modifiability were identified as well. They are shown as *hurt*-contribution links. Modifications in the software architecture can have a negative influence on the minimality of the trusted computing base and vice versa. The other security principles are affected by changes as well. Tamperproofness can easily be breached if a modification is performed in a wrong way. Therefore, changes should only be made on those architectural parts that have not to be isolated due to security reasons.

These conflicts confirm our earlier assumption that security is in conflict with flexibility and scalability. However, at the principles level their interdependencies have been clarified and we have a much better understanding of the conflict than on the goal or the subgoal level. Anyway, the conflict between the fundamental security principles and the subgoal modifiability cannot be resolved in this transition of the Goal Solution Scheme. The conflict resolution has to be postponed to the next design step, when a related solution can be analyzed more precisely than the principles.

Decision about Solution Instruments

In the third design step the actual transformation of the non-functional properties to a functional solution is performed. This step is closely related to the third step of the QASAR method (Bosch, 2000). A similar mapping of solution instruments to goals can also be found in the NFR framework and the i* framework, where operationalizations, or tasks respectively, are assigned to decomposed softgoals.

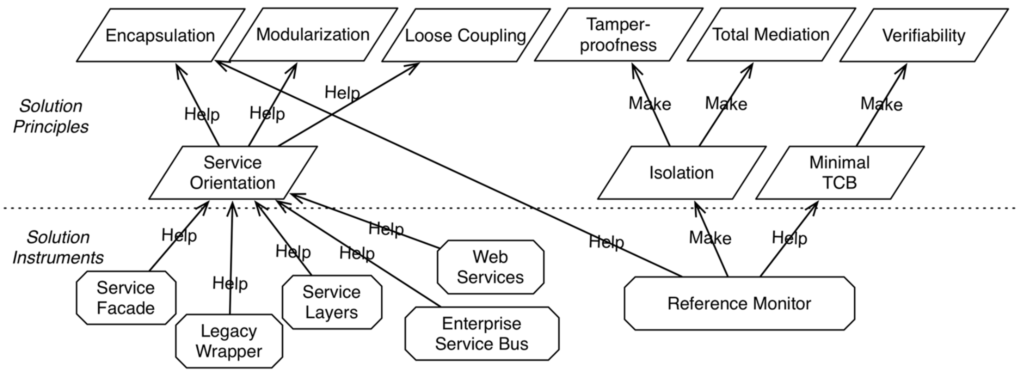
In our method solution instruments can be functional concepts or even existing technical components, which either support the realization of non-functional goals or completely fulfill some of them. In this third design step a large number of solution instruments is possible. In order to find the most adequate ones, the designer weights the different alternatives, akin to the last step.

The explicit linkage from goals to principles and solution instruments classifies the latter ones according to their contributions to the non-functional goals. The Goal Solution Scheme serves as a knowledgebase, which enables the incremental collection and the reuse of the solutions in a goal-oriented way.

Of course, the decisions are also influenced by other technical or organizational requirements and constraints. For example the technical component JGoodies (2008) explicitly facilitates usability with its subgoal user satisfaction by an easy alignment and balancing of visual elements. However, it cannot be chosen, if the project demands for the C++ programming language, because JGoodies is based on Java and Swing. To consider such architectural constraints, a two-stage process can be applied. In a first step, all solutions that are inappropriate are ignored. In a second step, the remaining ones are ranked according to their satisfaction of the goals to find the best solution.

Figure 6 shows a part of the Goal Solution Scheme for the transition from principles to solution instruments for our case study. To realize service orientation, and thus the principles encap-

Figure 6. Cut-out of the transition from principles to solutions for the case study



sulation, modularization, and loose coupling, the solution instruments *Web Services* and *Enterprise Service Bus* (ESB) for the integration of the MES and ERP are chosen. The reason for these decisions is that well-defined web services according to the Service-Oriented Architecture (SOA) paradigm inherently reinforce those principles (Erl, 2007). A component-based Common Object Request Broker Architecture (CORBA), for example, could have been an alternative for a service-oriented architecture. However, for our case study a CORBA infrastructure was not available.

As an example from our case study, one realized service shall be mentioned. The service *MachineAvailability* can be used for the interaction of the detailed planning of an MES and the general planning of an ERP. Using this service the ERP can request status information about machines, such as their availability. When implementing the web services, architectural and design patterns, such as *Service Layers* (Fowler, 2003; Erl, 2008), *Service Facade*, or *Legacy Wrapper* (Erl, 2008), contribute to the realization of the principles, and hence, to the non-functional goals.

The application of a reference monitor solves the integration of the security aspects. The security principles are integrated with the help of the ESB to gain control of the communication between the MES and ERP system and to isolate the security-relevant architectural parts. Aside, it must be considered to keep the TCB as small as possible. A discussion on the integration of security with web services can be found in (Fischer & Kühnhauser, 2008). With this kind of solution the conflict between the security principles and the subgoal modifiability, which was detected in the last design step, cannot be resolved completely. However, it can be implemented in a controlled way by controlling access to the security relevant functionality. Hence, the realization of a reference monitor not only positively contributes to the reference monitor principles, but also helps encapsulation, and therefore, even modifiability despite the conflicts.

As an alternative to the reference monitor, in an ad-hoc approach or according to the discussion by Chung et al. (2000), one could have considered only multiple single solution instruments, such as encryption mechanisms, or roles and rights, for security purposes. Of course, these solution instruments can contribute to confidentiality and maybe availability and integrity. However, as a drawback, without considering the reference monitor principles the system would be much more vulnerable.

In this step of the TraGoSoMa design method again all decisions about solution instruments for the solution principles are made explicit by traceability links. As shown in Figure 6 all solution instruments are mapped to the corresponding solution principles. The chosen solution instruments, such as the patterns Service Layers, Service

Facade, and Legacy Wrapper, are traced with links of the type realize. Additionally, elaborated alternatives not discussed here can also be linked with the type contribution and may be reused later. Figure 6 depicts only the mentioned solution instruments. Actually, much more solution instruments are contained, and the architect can easily extend them by additional ones.

Merging the Functional Solutions

In the fourth step of TraGoSoMa's second phase the solution instruments from both origins have to be merged, from functional goals and from non-functional ones. In this step a balancing between both types of requirements has to be performed (Harrison & Avgeriou, 2007). The functional requirements—the first type—have been elaborated into candidates for functional components as in phase 1 (see Figure 1) by a functional decomposition, for example following the FAD method by QASAR. For non-functional goals—the second type—solution instruments in the form of components are integrated, which are developed according to the Goal Solution Scheme in the second phase.

The merge can be performed by architectural change operations of different types. The simplest case is to only *add* the functional components that implement non-functional requirements from the second phase to the components of the first phase. A second type of transformation is to *replace* functional components. For example, if a functional component from phase one is insufficient in fulfilling the non-functional requirements, it is replaced by the solution instruments elaborated in the second phase. A third type of transformation is to *remove* a previously decomposed functional component to enable the realization of a non-functional goal. Baldwin and Clark (2000) mention three more elementary types of changes called modular operators: *split* a component into two, *port* a component for extraction to a new one, and *inversion of hierarchy* for moving components from a lower position in a hierarchy to a higher one. As another example, we mention the implementation of security goals. To solve this task by the reference monitor concept (Anderson, 1972) a separation of security-relevant and security-irrelevant functional components is performed to achieve a minimal Trusted Computing Base applying the *inversion of hierarchy* operation (for more details see Bode et al., 2009). In the case security-relevant and security-irrelevant functions are covered by one component, it has to be *split* or partly *ported*.

As the result of the merge a software architecture has been developed. Its components are functional ones, which can be implemented directly. All responsibilities that are due to functional and non-functional requirements are assigned to these components.

Evaluation of the Decisions

For an early assessment of the architectural design any iteration should include an evaluation. The assessment technique depends on the characteristics of the assessed products and on the criteria to be evaluated.

There are several well-established assessment techniques for software architectures. Two types are especially suitable—questioning and measuring techniques. The techniques of the first type, for example structured scenario-based inspections with the Architecture Tradeoff Analysis Method (ATAM) (Kazman et al., 2000), are performed by experts. Based on the assessment criteria, the scenarios are established, for example an intrusion scenario for an evaluation regarding security or an extension scenario regarding maintainability. Evaluations of this type can be performed early, even if the architecture is not complete. Performing the evaluation in a structured, formally defined way by external experts can reduce the disadvantage of the subjective nature of the result.

The measuring techniques provide objective, quantitative results. However, they require formal models and well-defined evaluation criteria. Examples for this type are metrics for architectural quality, e.g., for modularity by relating the number of all inner dependencies of a component to the outer ones. In (Brcina & Riebisch, 2008) traceability links between model elements are evaluated for evolvability, and for example the effects of tangled or scattered components are assessed as criteria for architectural quality.

FUTURE RESEARCH DIRECTIONS

The consideration of non-functional properties constitutes a long-term goal for architectural development, even beyond SOA. Further work is needed to close the gap between requirements engineering and architectural design regarding integration of methods. Maybe aspect-oriented techniques can help regarding this issue as well. Moreover, there is a special need for integrating the mentioned concepts with the Model-Driven Architecture (MDA) approach.

The need for a rigor specification of non-functional requirements can be fulfilled only for some categories, e.g., security. Many others, e.g., usability, flexibility, and scalability, are specified by informal or semi-formal descriptions. Ontologies can help to analyze the semantics of these descriptions based on the terms and their relations.

For the early evaluation of the results of architectural decisions, a prototyping is necessary in some cases, for example for requirements regarding efficiency and scalability. The generation of the necessary prototypes shall be based on the architectural decisions and documents to minimize the additional effort.

CONCLUSION

In this paper the architectural design method TraGoSoMa has been presented, which provides improved support and guidance for the architect to consider non-functional properties. The method aims at the definition of components and their implementation illustrated with a SOA example. As an important element the method introduces the Goal Solution Scheme, which represents the alignment between non-functional goals, their subgoals, the applied solution principles, and the solution instruments for implementing the required properties. By using this scheme, the conflict resolution between competing goals is supported, the solution principles and the solution instruments are classified according to the non-functional goals supported by them, and the systematic selection of solution instruments during architectural decisions is facilitated. Furthermore, the Goal Solution Scheme supports the accumulation of architectural know-how by a stepwise extension of the classification and the solution instruments. The scheme expresses traceability links in an explicit way, while the method facilitates their establishment, and thus, supports the maintainability of the software. Especially important is the traceability between non-functional goals, such as scalability, efficiency, and security, and the chosen solution instruments.

Regarding state-of-the-art methods, the contributions of the method can be compared to the QASAR method as well as to the works of Chung et al. and Yu as discussed in the background section. The QASAR method is extended by enhancing its second design phase by a systematic procedure for non-functional properties, which reduces the effort for transformations, and by improving the third phase by reducing the scattering of the changes to the design. In comparison to the goal-oriented approaches from requirements engineering, we consider architectural constraints from the environment as well as interdependencies with technical solution instruments. Furthermore, as a novelty step, we map non-functional goals to solution principles and integrate the goal-oriented activities in the context of an architectural design method. We use a design decision process, especially considering well-known solution principles to find conflicts, synergies, and solution instruments. Beyond, we include the establishment of traceability links into our method.

The method is applied for the redesign of a Manufacturing Execution System (MES) in order to integrate it into a service-oriented environment of business systems in manufacturing. With the help of this case study, we illustrate the several design steps, and therefore, show the importance of such a methodical design and its relevance for industrial use.

REFERENCES

Amyot, D. (2003). Introduction to the user requirements notation: Learning by example. *Computer Networks*, *42*(3), 285–301. doi:10.1016/S1389-1286(03)00244-5

Anderson, J. P. (1972). *Computer security technology planning study* (Tech. Rep. ESD-TR-73-51), L. G. Hanscom Field, Bedford, MA, USA: U.S. Air Force, Electronic Systems Division, Deputy for Command and Management Systems, HQ Electronic Systems Division (AFSC).

Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity* (*Vol. 1*). Cambridge, MA: MIT Press.

Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). The goal question metric approach. In Marciniak, J. (Ed.), *Encyclopedia of software engineering*. Wiley.

Bass, L. J., Klein, M., & Bachmann, F. (2002). Quality attribute design primitives and the attribute driven design method. In F. van der Linden (Ed.) *Software product-family engineering, 4th International Workshop, PFE 2001, Revised Papers,* (pp. 169-186). Berlin: Springer.

Bengtsson, P., Lassing, N., Bosch, J., & van Vliet, H. (2004). Architecture-level modifiability analysis (ALMA). *Journal of Systems and Software*, *69*(1-2), 129–147. doi:10.1016/S0164-1212(03)00080-3

Bode, S., Fischer, A., Kühnhauser, W., & Riebisch, M. (2009). Software architectural design meets security engineering. In *Proceedings of the 16th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ECBS).* (pp. 109-118). USA: IEEE.

Bode, S., & Riebisch, R. (2009). Tracing quality-related design decisions in a category-driven software architecture. In Liggesmeyer, P., Engels, G., Münch, J., Dörr, J., & Riegel, N. (Eds.), *Proceedings of Software Engineering 2009* (pp. 87–98). Bonn, Germany: Köllen.

Bondi, A. B. (2000). Characteristics of scalability and their impact on performance. In *Proceedings of the 2nd International Workshop on Software and Performance (WOSP '00),* (pp. 195-203). New York: ACM.

Bosch, J. (2000). *Design and use of software architectures*. New York: Addison Wesley.

Brcina, R., & Riebisch, M. (2008). Architecting for evolvability by means of traceability and features. In *Proceedings of the 4th International ERCIM Workshop on Software Evolution and Evolvability (Evol'08) at the 23rd IEEE/ACM International Conference on Automated Software Engineering,* (pp. 235-244). USA: IEEE.

Chappel, D. A. (2004). *Enterprise service bus*. USA: O'Reilly Media.

Chung, L., Nixon, B. A., Yu, E., & Mylopoulus, J. (2000). *Non-functional requirements in software engineering*. Norwell, MA: Kluwer Academic Publishing.

Cleland-Huang, J., Settimi, R., BenKhadra, O., Berezhanskaya, E., & Christina, S. (2005). Goal-centric traceability for managing non-functional requirements. In *Proceedings 27th International Conference on Software Engineering,* (pp. 362-371). New York: ACM.

Clements, P., Bachman, F., Bass, L., Garlan, D., Ivers, J., & Little, R. (2003). *Documenting software architectures: Views and beyond*. Amsterdam: Addison-Wesley Longman.

Department of Defense (1985). *Trusted computer system evaluation criteria.* DoD 5200.28-STD.

Duboc, L., Rosenblum, D., & Wicks, T. (2006). A framework for modelling and analysis of software systems scalability. In *Proceedings of the 28th International Conference on Software Engineering ICSE '06,* (pp. 949-952). New York: ACM.

Duboc, L., Rosenblum, D., & Wicks, T. (2007). A framework for characterization and analysis of software system scalability. In *Proceedings of the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering ESEC-FSE '07,* (pp. 375-384). New York: ACM.

Dueñas, J. C., & Capilla, R. (2005). The decision view of software architecture. In *Proceedings of the 2nd European Workshop on Software Architecture* (LNCS 3527), (pp. 222-230). Berlin: Springer.

Eden, A., & Mens, T. (2006). Measuring software flexibility. *IEE Proceedings. Software*, *153*(3), 113–125. doi:10.1049/ip-sen:20050045

Egyed, A. (2001). A scenario-driven approach to traceability. *In Proceedings of the 23rd International Conference on Software Engineering ICSE'01,* (pp. 123-132). Washington, DC: IEEE.

Erl, T. (2007). *SOA: Principles of service design*. Upper Saddle River, NJ: Prentice Hall.

Erl, T. (2008). *SOA design patterns*. Upper Saddle River, NJ: Prentice Hall.

Fischer, A., & Kühnhauser, W. E. (2008). Integration von Sicherheitsmodellen in Web Services. In P. Horster (Ed.), *D.A.CH security 2008*. Hannover, Germany: eMedia.

Folmer, E., & Bosch, J. (2003). Usability patterns in software architecture. In *Proceedings of the 10th International Conference on Human-Computer Interaction HCII2003 vol. I,* (pp. 93-97).

Fowler, M. (2003). *Patterns of enterprise application architecture*. Boston: AddisonWesley.

Galster, M., Eberlein, A., & Moussavi, M. (2006). Transition from requirements to architecture: A review and future perspective. In *Proceedings Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06),* (pp. 9-16). USA: IEEE.

Gasser, M. (1988). *Building a secure computer system*. New York: Van Nostrand Reinhold Co.

Goguen, J. A., & Meseguer, J. (1982). Security policies and security models. In *Proceedings IEEE Symposium on Security and Privacy,* (pp. 11-20). Washington, DC: IEEE.

Gotel, O. C. Z., & Finkelstein, A. C. W. (1994). An analysis of the requirements traceability problem. In *Proceedings of the First International Conference on Requirements Engineering,* (pp. 94-101). USA: IEEE.

Grau, G., & Franch, X. (2007). A goal-oriented approach for the generation and evaluation of alternative architectures. In F. Oquendo (Ed.), *Software architecture, Proceedings First European Conference, ECSA 2007.* (pp. 139-155). Berlin: Springer.

Harrison, N., & Avgeriou, P. (2007). Pattern-driven architectural partitioning: Balancing functional and non-functional requirements. In *Proceedings Second International Conference on Digital Telecommunication (ICDT'07),* (pp. 21-26). USA: IEEE.

Hill, M. D. (1990). What is scalability? *SIGARCH Computer Architecture News*, *18*(4), 18–21. doi:10.1145/121973.121975

Hofmeister, C., Nord, R., & Soni, D. (2000). *Applied software architecture*. New York: Addison Wesley.

ISA. (2000). *ISA–95.00.01–2000 Enterprise-control system integration. Part 1: Models and terminology*. North Carolina: ISA.

ISO. (2001). ISO/IEC 9126-1 International standard. Software engineering–product quality –part 1: Quality models.

ISO. (2005). ISO/IEC 27001:2005 Information technology–security techniques–information security management systems–requirements.

JGoodies. (2008). *JGoodies: Java user interface design.* Retrieved October 13, 2008, from http://www.jgoodies.com/

Jiang, H., Nguyen, T. N., Chang, C. K., & Dong, F. (2007). Traceability link evolution management with incremental semantic indexing. In *Proceedings 31st Annual International Computer Software and Applications Conference (COMPSAC 2007),* (pp. 309-316). USA: IEEE.

Kazman, R., Klein, M., & Clements, P. (2000). *ATAM: Method for architecture evaluation.* (Tech. Rep. CMU/SEI-2000-TR-004). Pittsburgh: Carnegie-Mellon University, Software Engineering Institute.

Lampson, B., Abadi, M., Burrows, M., & Wobber, E. (1992). Authentication in distributed systems: Theory and practice. *ACM Transactions on Computer Systems*, *10*(4), 265–310. doi:10.1145/138873.138874

Letelier, P. (2002). A framework for requirements traceability in UML-based projects. In *1st Int. Workshop on Traceability in Emerging Forms of SE (TEFSE'02),* (pp. 32-41). Edinburgh, UK.

Liu, L., & Yu, E. (2001). From requirements to architectural design–using goals and scenarios. In *From Software Requirements to Architectures Workshop (STRAW 2001),* (pp. 22-30). Toronto, Canada.

Mäder, P., Gotel, O., & Philippow, I. (2008). Rule-based maintenance of post-requirements traceability relations. In *Proceedings of the 2008 16th IEEE International Requirements Engineering Conference (RE '08),* (pp. 23-32). USA: IEEE.

Mäder, P., Philippow, I., & Riebisch, M. (2007). Customizing traceability links for the unified process. In *Proceedings of the Third International Conference on the Quality of Software-Architectures (QOSA2007)* (LNCS 4880). (pp. 47-64). Berlin: Springer.

Mäder, P., Riebisch, M., & Philippow, I. (2006). Traceability for managing evolutionary change. In *Proceedings of the 15th International Conference on Software Engineering and Data Engineering (SEDE-2006),* (pp. 1-8). USA: ISCA.

Marinescu, R., & Ratiu, D. (2004). Quantifying the quality of object-oriented design: The factor-strategy model. In *Proceedings 11th Working Conference on Reverse Engineering (WCRE 2004),* (pp. 192-201). USA: IEEE.

Morgan, G. (2006). *Design for flexibility.* Retrieved October 13, 2008, from http://blogs.msdn.com/gabriel_morgan/archive/2006/10/03/Design-for-Flexibility.aspx

Nelson, K., Nelson, H., & Ghods, M. (1997). Technology flexibility: Conceptualization, validation, and measurement. In *Proceedings of the Thirtieth Hawaii International Conference on System Sciences, Vol. 3,* (pp. 76-87). Washington, DC: IEEE.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Ollson, T., & Grundy, J. (2002). Supporting traceability and inconsistency management between software artifacts. In *Proceedings of IASTED International Conference on Software Engineering and Application*.

Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM*, *15*(12), 1053–1058. doi:10.1145/361598.361623

Pinheiro, F. A. C. (2004). Requirements traceability. In Leite, J. C. S. P., & Doorn, J. (Eds.), *Perspectives on software requirements* (pp. 91–113). Norwell, MA: Kluwer Academic Publishers.

Pohl, K. (1996). PRO-ART: Enabling requirements pre-traceability. In *Proceedings of the Second International Conference on Requirements Engineering ICRE'96,* (pp. 76-84). Washington, DC: IEEE.

Ramesh, B., & Jarke, M. (2001). Toward reference models for requirements traceability. *IEEE Transactions on Software Engineering*, *27*(1), 58–93. doi:10.1109/32.895989

Roy, J., Kealey, J., & Amyot, D. (2006). Towards integrated tool support for the user requirements notation. In R. Gotzhein & R. Reed (Eds.), *System analysis and modeling: Language profiles.* (pp. 198-215). Fifth International Workshop, SAM 2006. Berlin: Springer.

Shneiderman, B. (1992). *Designing the user interface: Strategies for effective human-computer interaction* (2nd ed.). Boston: Addison-Wesley.

Siedersleben, J. (2004). *Moderne Software Architektur: Umsichtig planen, robust bauen mit Quasar.* Heidelberg, Germany: dpunkt.verlag.

Standards Coordinating Comittee of the Computer Society of the IEEE. (1990). *IEEE standard glossary of software engineering terminology.* IEEE Std 610.12-1990.

VDI. (2007). *VDI 5600: Fertigungsmanagementsysteme. Manufacturing Execution Systems (MES)*. Berlin: Beuth.

Winkler, S., & von Pilgrim, J. (2010). A survey of traceability in requirements engineering and model-driven development. In *Software and Systems Modeling*, 9(4), 529-565.

Wohlfarth, S. (2008). *Entwicklung eines rationalen Entscheidungsprozesses für Architekturentscheidungen.* Unpublished doctoral dissertation, Ilmenau University of Technology, Germany.

Yu, E. (1995). *Modelling strategic relationships for process reengineering.* Unpublished doctoral dissertation, Dept. of Computer Science, University of Toronto, Ontario, Canada.

Zeng, D., & Zhao, J. (2002). Achieving software flexibility via intelligent workflow techniques. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences, HICSS,* (pp. 606-615). Washington, DC: IEEE.

KEY TERMS AND DEFINITIONS

Architectural Design Method: A systematic approach with analysis, synthesis, and evaluation activities to create a software architectural description from functional as well as non-functional requirements taking organizational and technological constraints into account.

Enterprise Resource Planning: The complex task to efficiently use all resources of an enterprise, such as financial assets, production facilities, or personnel, to control and optimize business processes.

Manufacturing Execution System: A software system responsible for the organization and execution of the production process in a factory with numeral scopes of duty as management of all required activities within the production process or the exchange of information with the environment as to ERP systems (for a comprehensive definition see (VDI, 2007) for a comprehensive definition).

Non-Functional Requirement: Also quality requirement, quality attribute, or quality goal – a software requirement that describes not what a software system has to do, but how it should be done and under which constraints, and therefore, defines its quality (for a comprehensive discussion see (Chung et al., 2000) for a comprehensive discussion).

Software Architecture: The description of the organizational structure of a software system, its architectural elements, their properties, interfaces, relations, and behavior, as well as a set of decisions and guidelines for the design of the system.

Software Quality: The totality of characteristics of a software product that bear on its ability to satisfy specified requirements (cf. ISO, 2001)).

Traceability: The capability to track and recover in both a forwards and backwards direction the development steps of a software system and the design decisions made during on-going refinement and iteration in all development phases by relating the resulting artifacts of each development step to each other (based on (Gotel & Finkelstein, 1994)).

This work was previously published in Non-Functional Properties in Service Oriented Architecture: Requirements, Models and Methods, edited by Nikola Milanovic, pp. 1-23, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Section 2
Development and Design Methodologies

This section provides in-depth coverage of conceptual architecture frameworks to provide the reader with a comprehensive understanding of the emerging developments within the field of Industrial Engineering. Research fundamentals imperative to the understanding of developmental processes within Industrial Engineering are offered. From broad examinations to specific discussions on methodology, the research found within this section spans the discipline while offering detailed, specific discussions. From basic designs to abstract development, these chapters serve to expand the reaches of development and design technologies within the Industrial Engineering community. This section includes 14 contributions from researchers throughout the world on the topic of Industrial Engineering.

Chapter 13
Learning Parametric Designing

Marc Aurel Schnabel
The Chinese University of Hong Kong, Hong Kong

ABSTRACT

Parametric designing, its instruments, and techniques move architectural design education towards novel avenues of deep learning. Akin to learning and working environments of engineering and manufacturing, it offers similar advantages for architects. Yet it is not as simple as using another tool; parametric designing fundamentally shifts the engagement with the design problem. Parametric designing allows architects to be substantially deeper involved in the overall design and development process extending it effectively beyond production and lifecycle. Leaning parametric design strategies enhance architects' critical engagement with their designs and their communication. Subsequently, the computational aid of parametric modelling alters substantially how and what students learn and architects practice.

INTRODUCTION

Parametric design techniques offer obvious advantages for engineering and manufacturing processes, now architects have emerged to apply these methods in their working environment suggesting solutions and novel designs at an earlier stage of the process. Through the coupling of architectural design with parametric modelling methods, the chapter presents techniques that enhance students' learning and knowledge about designing and architectural building processes. This allows a deeper comprehension of the design objectives and aids architectural designers in their decisions to find solutions.

A dilemma of semester-based teaching is that students reach their highest level of skills and experience at the end of a term, after which they leave for their break and are therefore unable to apply their freshly gained knowledge immediately. At the beginning of the next following term, however, the knowledge and skills they had gained earlier are likely to be either inactive or not employed, and learning foci may have shifted to other aims.

DOI: 10.4018/978-1-4666-1945-6.ch013

The architectural design studio presented here addressed these issues by integrating the learning experience from the beginning by focusing on parameters that create or inform about the design. The objective of this 'parametric designing' was to allow students to understand the impact each step and variable has on the design and follow the impact it has onto the project. Students developed and communicated their design parameters by utilizing their knowledge throughout the design-studio environment. Because of this, students began to think about design problems in different ways. The studio explored design by basing it on parameters and their connecting rules. In order to build up a philosophy around parametric dependencies and relationships, the participants used digital instruments that aided them to create and express their designs. With these instruments, they could develop expertise to engage creatively in designing. The studio cumulated in an architectural art exhibition highlighting the coupling of architectural design with digital modelling and fabrication methods. Students presented architectural solutions that challenged and addressed environmental and programmatic issues, dimension, space and volume, as well as theoretical and conservational topics, resulting in novel designs created with freedom of innovation, interpretation, and definition some of which without any boundaries. The notion of non-conformity added to the core of this collection of works, held together by the idea of spatial concepts and parametric designing in architecture.

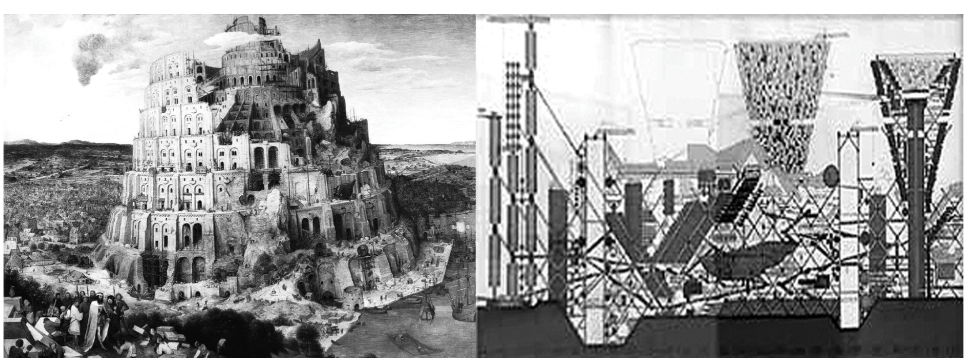
BACKGROUND

Pieter Bruegel, a Netherlands' Renaissance painter, depicted a representation of the 'Tower of Babel' as a building that is constantly redefining its needs, as it grows larger and more complex (Figure 1, left). The painting shows a tower nearly reaching the clouds and illustrates all the problems then associated with cities, buildings and life within and the constant change and reaction to new situations during the process of building.

The exploration of the relationship between human beings and the natural world, and the subsequent implications of interactions between them, has deep roots in our social and cultural understanding of society. Cities, therefore, are direct reflections of their inhabitants, as their architectural expressions directly influence the living conditions of their people. In recent practice, architects have designed and described buildings through the means of (master-) plans, sections, elevations, or descriptions of render-perfect, complete architectures in which change was not part of the picture. A few, however, have tried different approaches to communicate architecture.

In the 1960s and early 1970s, Archigram already presented an idea that reacted against the

Figure 1. Right: Pieter Bruegel's 'Tower of Babel. Right: Archigram's Plug-in City



permanence of houses in what it called the "Plug-in City" (Figure 1, right). They proposed architecture that is ever changing and adaptable to different social and economic conditions (Karakiewicz, 2004). Their proposal did not develop further than a conceptual stage, yet it lays in contrasts to the common practice that also Le Corbusier describes as non-intelligent building machines, whereby these machines would not think, and would therefore be unable to adapt to change.

More recently 'LAB Architecture Studio' translated planning codes of Beijing's 'Soho Shang-Du' into series of parametric design rules whereby the outcome both complies with and confounds the rigid regulations (Davidson, 2006). As a result, the architects did not prescribe a fixed definition of architecture, but a set of rules and instructions that inform about and can generate the outcome. This allows a reaction on a variety of site-specific variables that can be modified according to the need, yet fit into the overall design intents of the architects.

These samples point out the constant demand for architecture to adapt and react to a variety of parameters that are driven by its use and context. The gap between the architectural design conceptions and the translation of these designs into the real built environment can be addressed fundamentally differently by an intersection of process and outcome (Eastman, 2004). Parametric design and digital fabrication techniques suggest controllable and adaptable solutions at an earlier stage of the process that react to the given situations and the outcomes.

PARAMETRIC DESIGN STUDIO

Architectural design studios are an essential learning experience for architectural students. Their traditions and proceedings are well established. Studios go beyond pure skill training and require reflection upon, and the creation of, knowledge. These studios are, additionally, informed and supplemented by courses and seminars that contribute to the overall learning goals. Yet there can be a gap between skill training and application of knowledge. At the end of the studio, students may not be able to identify how they arrived to their solution and solves a given problem, or what were the individual contributors that made their design successful.

In computational architectural studios, the same phenomena can be observed. These studios present the underlying concepts of architectural design using computational instruments, and have at the same time to provide for software skills and other technical knowledge (Kvan, 2004A). The integration of digital media into design studio curricula often fails, because the compound acquisition of skills prevents a deep exploration of design and the theoretical aspects involved at the same time. Participants can employ computational instruments within a studio context only after they have learned subject matters and acquired proficiency in their skills. By then, the studio may consider these skills no longer valid or has ended.

Parametric applications have inherited two crucial elements. These are that all entities start in a multi-dimensional space and allow the study of architectural conditions in a cloud of data and variety of representations, rather than the conventional two-dimensional or layered design environment. The underlying notion of parametric designing is based on the contextual construction of a formal and spatial systemic intelligent simulation; or in other words data, variables, and their relationship to other entities, which can then respond to variations of necessities (Ambrose, 2009). Students learn about cause and effect in both abstract environments as well as at specific situations of their design task. This is where architectural education is in the process of changing fundamentally. Design studios and courses are now increasingly reacting to the quantum leap architectural computing has presented to design education, and introducing computational parametric tools to the design studios that go beyond

Computer Aided Design (CAD) (Picon, 2010). Yet one has to be careful that novel technologies and learning methodologies offer current pedagogies to address certain known issues and cannot eliminates all problems connected with learning and education.

PARAMETRIC DESIGNING

Architecture in general can be expressed and specified in a variety of ways. Commonly, drawings describe geometric properties that can explain, depict, and guide the construction of buildings or streets. Alternatively, performance specifications can describe observed behaviours. It is also possible to describe properties as relationships between entities. Spreadsheets, for instance, specify the value of each cell as the result of calculations involving other cell entries. These calculations or descriptions do not have to be explicit. Responsive materials change their properties in reaction to the conditions around them. A thermostat senses air temperature and controls the flow of electric current, and hence the temperature of the air supplied. Using such techniques, artists have created reactive sculptures and architects have made sentient spaces that react to their occupants or other relevant factors. Streetlights turn on if light levels fall below a threshold; traffic flow can be regulated according to need; walls can move as users change location.

Links to a variety of data can be established and subsequently serve as the bases to generate geometric forms using parametric design instruments. When designing spaces, it is usual to collect some data of the type of architectural qualities desired. These are then, for example, translated into master plans, which are themselves specific spatial descriptions. Performance requirements for spaces can then be written, linking the description of the architecture to experiential, financial, environmental, or other factors (Picon, 1997).

Design studios mimic the typical working processes of the architectural profession and are the essential learning experience for architectural students. Research is now looking into how the framing of design problems using parametric methods enhances the overall process (Schnabel et al., 2004). The here presented studio, therefore, couples parametric methodologies within the generation of architectural design, ultimately reframing the problem and proposing new answers to design thinking and learning. Participants in this study solved a typical architectural design problem using computational applications that focused on the parametric dependencies of spatial entities, generative scripts, and form finding. The re-representation of the design intent sharpens the question at its centre (Gao and Kvan, 2004), while taking full advantage of available parametric modelling software to explore it. This approach tested the limitations set by conventional, design-only methods. The cognitive aspects of the design generation and their relationships to parametric design methods operated as an influential factor for the understanding of the projection of design intent, framing, generation of spatial knowledge within architectural design and the reflection about the outcome produced in this process (Ambrose, 2009).

Problem Framing

The studio engaged the participants in design processes by using sets of variables and series of relations to question, create, and define the form and function of the resulting designs. Thus, the students examined interaction techniques between their design intent, their framing of the design problem, their subsequent generation and reflection on their development by testing the rules and parameters. Participants engaged in a collaborative architectural design studio involving the creation and fabrication of architecture. This formed the basis for a transfer of knowledge to the larger context of the issues ahead in their future professional careers (Riese and Simmons, 2004).

The studio took a distinctive neighbourhood within the larger urban context of Sydney, Australia as its base of exploration. The specific site surrounding, a mix of residential, public and commercial buildings, offered a medium dense area with a variety of architectural languages.

Driven by a fast growing population, an architectural strategy that steers further development was sought. The city's scale, its growth through migration and the need for new housing have an impact on its inhabitants' sense of place and sense of community. Earlier urban planning did not anticipate the changes that arose over years of population growth. Hence, a new strategy for development that could address these issues was sought to create a new identity for the place and the city itself (Forrest et al., 2002).

The site of the studio had typical architectural characteristics and requirements. Located at a riverbank, in close proximity of a parkland, cultural-, office- and residential buildings, the site offered a variety of inspiration as well as constrains for an architectural design exercise. Students had to address and responded to the local and overall conditions of open space, city, work, living and environment.

The studio built upon design studios where participants explored design methods and tools beyond their original definitions and perceived limits (Schnabel et al., 2004). To allow the students both to acquire skills and training within their studio and to apply this knowledge to their design, the studio had an integrated digital media component that addressed parametric modelling in architectural design.

Two groups of fifteen students of the postgraduate architectural program each joined this studio, which was guided by two design tutors and one architectural consultant in digital media. The studio was structured in four phases that related to and built upon each other (Figure 2). The aim was to acquire and integrate parametric design knowledge and to use it as the base of the design creation of their architectural proposal. As a result, the final design could be modified and manipulated based on the parameters and their dependencies, allowing the students to gain a deeper understanding of their design processes and outcomes as well as the reaction of their proposals with the various influences of the site.

Creating

The project's first component included the collection and understanding of data that arrived from the site. In order not to overwhelm the students, the tutors asked them during this first stage to limit them to investigating only two points of interests, which became their key-parameters. Hereby the students could focus on the selections of parameters that they believed would influence their building proposal or their site's perception, this parameter could be a real or abstract item (Figure 3).

The parameters they chose informed them about the variables and correlation of their guid-

Figure 2. Four phases and exhibition of the design studio with learning reflections and projections

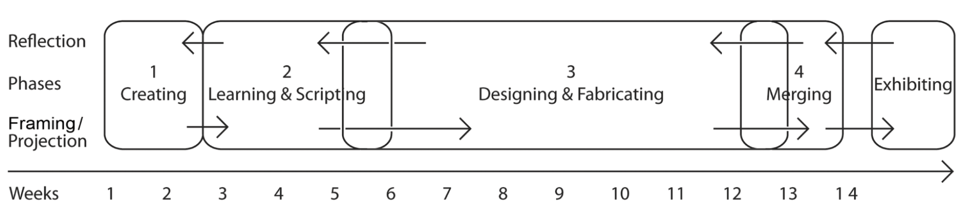
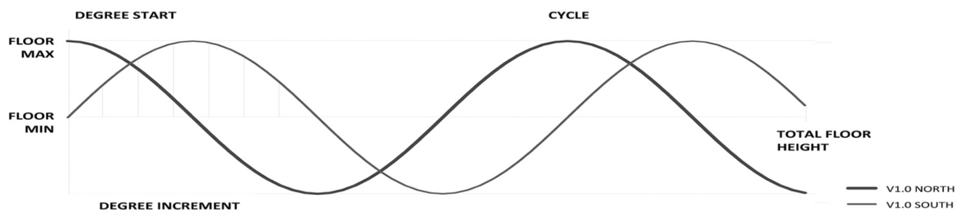


Figure 3. Two parameters (floor-heights) used by a student team (R Beson & N Minasian)



ing design principles that formed their initial rules. These provided them a description based on dependencies and interconnected relationships of relevant information. The chosen parameters helped the students to understand what impact certain variables may have on a design strategy and the design itself. This component concluded after two weeks with presentations of data, parameters, and individual interpretations of the site.

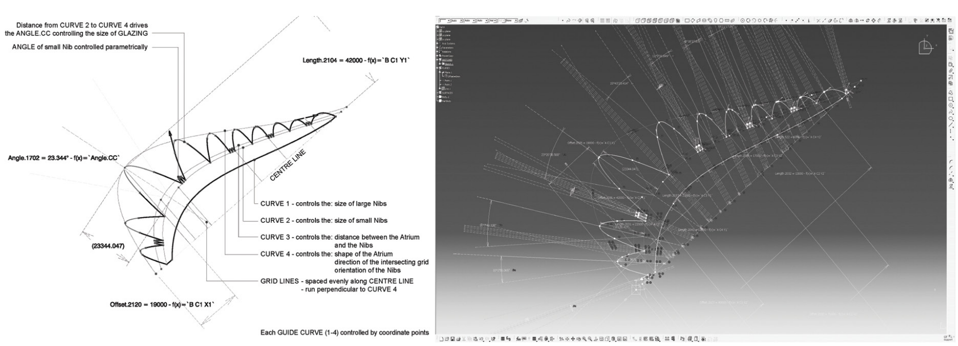
Learning

The program's second component focused on the understanding and creation of parametric concepts and the acquisition of design-application skills that allow rule-based three-dimensional design. Participants were trained intensively during studio time in the use of *Digital Project* TM (2004). This software allows users to not only create three-dimensional models, but also to establish rules, create parameters and their dependencies on a variety of entities (Figure 4).

Parametric functions require a different understanding of the conceptual approaches to design. Creating rules and dependencies, which then create the design, involved the students in a higher level of problem framing and definition of the concept of design. It allowed the visualization and modelling of highly complex forms that may result from non-traditional design data, such as noise data or spatial requirements.

The students focused on their own parametric and rule-based design analyses from the first component and subsequently studied mainly only the aspects relevant to these in relation to the use and operation of the software, the creation of rules, and parametric and generative design. During this phase, they used the time allocated to the design studio to establish a basic understanding of the software in its relationship to the design intent developed during the first phase. After three weeks of intensive training in architectural computing, the students reached a sufficient level of skills

Figure 4. Development of parameters, dependencies and rules within software analogue (left) and digitally (right) (Beson & Minasian)



that enabled them to use the parametric software as an aid for the creation of their own designs.

Scripting

'Script' is derived from the term for written dialogue in the performing arts, where actors are given directions to perform or interpret. Subsequently, 'scripting' is a creative process that describes the artistic intent of the designer. Scripts can define a set of rules that combines parameters in the named way. Software applications can be programmed and adjusted by scripts allowing for example repetitive tasks to be automated or to generate solutions that fit to a range of parameters (Biloria et al., 2005). Instead of using only compositional methods for designing, the students utilized scripts to form their own generative properties and base for their design exploration (Figure 5). Sourcing related or suitable general available scripts students quickly learned how to edit and control their design by amending the parameters or rules to fit their design intent (Celani, 2008). This phase differs greatly from conventional studios because students are engaging in software training and skill acquisitions of how to generate and manipulate instructions for computer programmes that can aid their design process.

Designing

The program's third component, scheduled for seven weeks, concentrated on design creation, reflection, and the communication of architectural design proposals. Using the data of the first component and the skills of the second, the students then started to establish and visualize their designs in three-dimensional forms that created spatial expressions of their findings and explorations.

Due to the emphasis on parameters, the studio was in particular interested in describing a building form by creating dependencies of parameters that defined the relationship of data to architectural expressions. With the use of a parametric modeller, it was easy to create geometric entities, solids and voids, and relate them to the context of the design task. This method made it obvious how one can learn about design and understand the various steps and elements through the logical steps laid out by the chosen parameters, variables, rules or scripts.

Fabricating

Another stage in the creative process is the fabrication of the digitally created designs. Recent computational applications and digital fabrication technologies have allowed architecture to take novel directions. The combination of architectural computation with computer-controlled machinery has nearly made it possible for shapes, however complex or irregular they seem to be, to be rationalized and created as physical entities with the ultimate aim to result in a buildable architecture (Oxman and Oxman, 2010). The studio subsequently made extensive use at all stages to explore

Figure 5. Variations generated by a script to modify facade-tiles (Beson & Minasian).

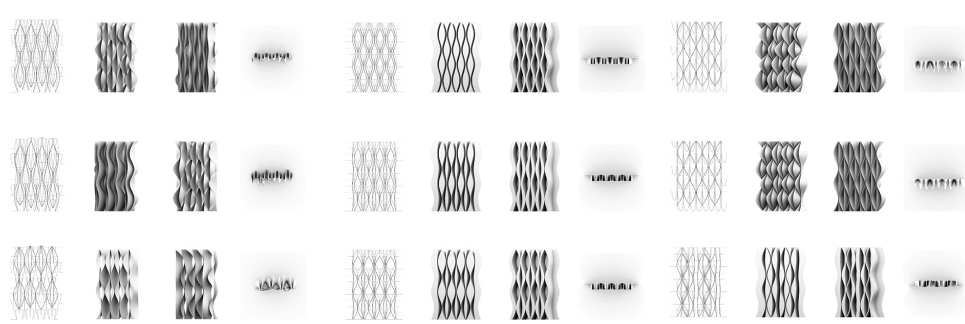
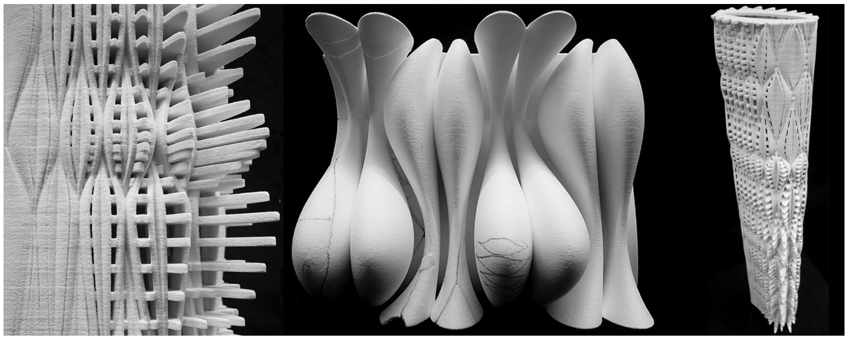


Figure 6. Facade details from the digital model fabricated by rapid-prototyping (Beson & Minasian)



the transformation of virtual design conceptions to physical objects via the use of computer-aided manufacturing (Figure 6).

Merging

The program's next following component brought together the various aspects and results of the earlier modules. Within two weeks, the students merged their individual designs into larger cluster files. This synthesis created compound descriptions and dependencies that were highly complex and interrelated, yet both the content as well as the tool allowed seamless communication to a larger audience by describing the rules and parameters (Figure 7). This phase created a design with shared authorship of all participants and allowed the students to study and understand the complexity and the interrelationships of architectural designing that they normally would have been unable to perceive immediately. Through their collaboration and exchange the students built up a collective intelligence that was driven by the individual contributions. The change of a single variable modified the whole design. Participants understood therefore the complex dependencies that one variable has in a large building and the impact it can have on the design.

Exhibiting

The design explorations culminated in an exhibition displaying the designers' engagement with parametric designing and fabrication (Figure 8). To mark the distinctive final stage in a celebrative conclusion of design development, the event exemplified how digital architectural design can conceptually and artistically engage with a

Figure 7. A joint model that combined facade, interior spaces and atrium details

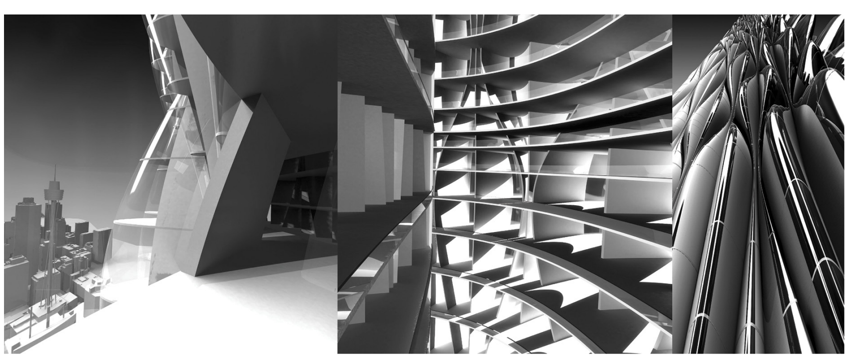
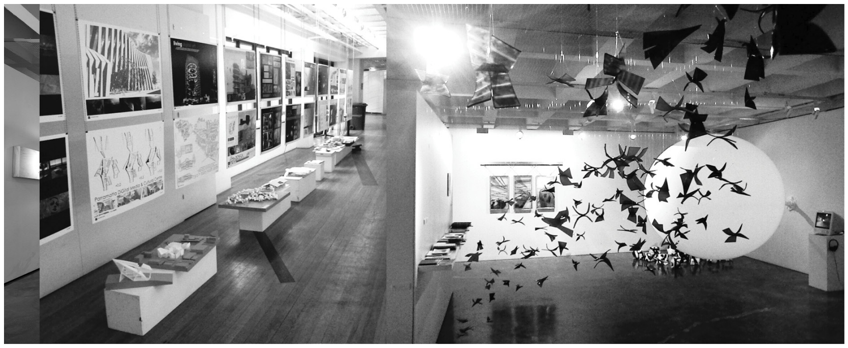


Figure 8. Exhibition of final designs at Brand Smart Centre (left) & Tin Shed Gallery, Sydney (right)



particular site, where a variety of solutions to problems in architectural design were developed from a diversity of multi-faceted and eccentric approaches (Schnabel and Bowller, 2007). The participating designers pushed creativity to new boundaries in definition of their artwork and cultural contexts, setting the direction for poetic viewpoints on innovation in architecture and spatial design. The exhibition forms a crucial learning experience whereby both processes and outcomes are presented in a formal way that is self-explanatory to a wider audience.

The compiling of all projects into a single exhibition removed the designers from the context of individual ownership, providing them with the invaluable opportunity to reflect on both, their own and their colleagues' proposals as a coherent collection of contributions towards one common engagement of design.

OUTCOMES

Participants of the parametric design studio were able to employ digital media skills from very early on throughout the studio and expand on these with their understanding and communication of design issues from there.

The students had already acquired a very high level of skills in using a specific parametric instrument within the first half of the studio. This enabled them to employ the instrument as an amplifier to learn about their designs. Subsequently, they were not limited by their knowledge or level of skills in order to be able to express themselves. The students produced a variety of individual design proposals as well as one large design-cluster. They created rules, scripts and parameters that allowed complex and interrelating designs to emerge. These representations could not be generated or communicated using traditional architectural design methods or instruments.

For example, one proposal related street lighting, neon-signs and display-windows with human activity around the building site. These parameters provided the engaging surface for the building mass. Subsequently they controlled the use, orientation and appearance of the building. The author took references to Japanese inner cities, where innovative ways of spaces are created by the means of lights, advertising and projections. Void, volume and density is controlled and created by the rhythm and intensity of lights. The student transferred this concept into parameters, which redefined the spatial understanding of the site and used these variables to create an architectural proposal.

Other results used parameters that related to the relationships between people and attraction to spaces with responsive structures. Students created self-opening canopies that reacted to people, activities, ferry schedules, weather conditions and the possibilities to collect rainwater to provide a comfortable environment in all conditions.

One team reacted with different floor-heights to various needs of public and private programme of their building and related their spaces to vista and light penetration of their building (Figures 3-6). These explorations then were merged with parameters controlling the interior space, atrium and program to form an overall design of a mix use building (Figure 7).

In the studio's last component, all students presented in-depth clusters of multifaceted architectural design proposals for the site. They demonstrated a high level of thinking processes resulting in the generation of compound rules and dependencies that finally create the architectural design schemes. Each student contributed simultaneously to create a variety of design proposals. The participants gained a high level of expertise with digital parametric tools as part of their development at the studio, and used this knowledge to design parametrically. The outcome clearly showed that thinking, learning and creating within parametric designing requires a novel and deeper understanding of the overall design goal and its anticipated outcome. The studio also showed that a social engagement with team members created a common knowledge to which everyone not only contributed but also benefitted. Students subsequently build up a social intelligence that allowed them to address both skills and design problems.

Students reported of the step learning curve of understanding how parametric software is structured differently to conventional software. They reported that they could not just design intuitively as they would do in conventional studios and had to stick to the rigour process of the parametric design methodology. The skill training in the software and the translation of the design intent proved not always to be straight forward. Some students only gained the full understanding of the potential parametric modelling offers at the end of the studio where all solutions were presented. While others had difficulties in developing a logic string of design steps that relate to the parametric approach, they preferred or felt back to intuitive or conventional designing.

Parametric modelling subsequently does not solve all issues connected with design-learning. It allows however, an alignment of cause and effect and a reflection of the design intent, the process and outcome. This differs from conventional design studios where these dependencies only seldom can be established. The studio allowed participants to learn about designing and problem framing. They were able to theorize and reflect on design creation for this and other design tasks. Consequently students engaged in a deeper learning that allows them to transfer and adapt their knowledge to new situations. Results can be explored at: www.parramatta.tk and www.disparallelspaces.tk.

DISCUSSION

In the early stages of computational architecture, designing in layers was a popular enrichment to conventional designing because it allowed architects to deal with problems that are more complex, with each different layer playing a specific role. It singled out issues and allowed dealing with them one at a time. Items that are more complex were divided into separate issues and dealt with one by one. Parametric design opens up a novel set of opportunities. It enables architects to study causes of problems and their relationships to, and dependencies on, other elements directly within a three-dimensional environment.

This shift of design thinking and creation needs to be addressed in the teaching and learning of design. Additionally, parametric designing provide for unpredictable events in connection with an overall architectural framework. Architects and architecture itself can respond to unplanned changes and their resulting consequences. The outcomes of this design studio showed that parametric dependencies allowed for such a level of ambiguity that is desired and required in creative and learning environments.

One objective of the studio was to frame an intellectual question that created design descriptions based on rules, scripts and parameters. The more interesting outcomes resulted from the ability to redefine and reframe the problems themselves by stepping out of preconceptions based on experience and exploring sets of unpredictable answers and then reflecting back on the starting point. Hence, in certain ways, parametric designing act at a higher level of the problem framing. The establishment of meta-rules has instituted a form of problem framing that demands the reference of one problem or parameter with other ones.

The learning outcomes of the parametric design studio demonstrate how non-linear design processes led to architectural design understandings that differ from conventional approaches to design learning due to their different nature of design thinking, framing, creation, and intuition. Despite three-dimensional representations of an architectural space being only a medium aiding the understanding and communication of spatial arrangements, the designers' comprehension of complex spatial qualities was enhanced by the parametric design environment, partly due to the logic structure and dependencies of one step to the next. The steep learning curve and the time needed to adjust to the parametric and sometimes stringent or seemingly limiting methodology of parametric designing shows that conventional designing is the pre-dominant approach to design and deeply routed in the design-thinking of students. Yet despite these difficulties students unanimously reported that the here presented studio helped to understand how to design and they highly valued the approach to thinking about and executing designing.

The use of parametric instruments allowed all students to design within an environment based on rules and generative descriptions, amplifying their understandings of creative processes and their learning outcomes. Each designer bridged the rift between their knowledge and ambition, creating architectural designs and learning about the act of architectural designing.

NEXT STEPS

The increasing marginalization of architects in the building industries (Bennetts, 2008) suggests that professional and educational ideals and professional work are poorly aligned. Unlike other professions, architects are trained in a variety of fields of knowledge and skills that are not directly related to the daily routine of the architectural praxis. Subsequently architecture students have an increasing amount to learn following graduation. Architects have discovered how digital instruments alter any aspects of their routines of working. However, academic and educational environments are not able to follow in the same speed. Learning designing has shifted from the single learner to a collective engagement with a variety of learners, novices, experts and instruments that aid, analyze, generate, design and review. Less than a decade ago many schools of architecture did not allow students to deliver CAD drawings for design projects assuming that would limit the exploration and understanding of design. In fact, the early experiments in using the computer in the design process quite often failed only because of the restrictions of the available infrastructure, facilities and skills. Today, students are familiar with architectural computing even before they enter the university (Dokonal and Hirschberg, 2003).

Still many questions remain unanswered and new questions arise in the relationship between architectural design and architectural computing. Architectural design is both an imagination and the ability to convey this idea. The learning of architectural design has to make use of the advantages that complex architectural computing offers without loosing the qualities of the established conventional methods. The current

'Net-Generation' (Oblinger and Oblinger, 2005) of learners, who are more conversant in using computational instruments than their teachers, are changing the dynamics of architectural education. This is a challenge to established curricula and institutions.

The herein presented studio is a successful attempt to integrate architectural computing into the learning environment by aligning skills and knowledge of the students with the objective to generate knowledge about designing, computation, architecture and realization.

Akin to Maver's (1995) comments, parametric designing and for that matter, architectural computing is certainly far from being resolved and offering the perfect solution. As the needs, goals and problems are rapidly developing architectural design and its learning needs to facilitate the evolution and progress. Synergies between the different realms, media and technologies are constantly evolving and adjusted to foster the evolution of architectural praxis and the building industry (Eastman et al, 2008).

CONCLUSION

The parametric design studio method presented in this chapter addressed computational concepts of architectural designing that influence the recent learning environment of architectural education. It coupled the setting of studio-learning with an in-depth digital media training in order to close the gap between acquisition of skills and the reflection of knowledge, as well as to explore new avenues of framing and integrating compound design issues. The use of digital parametric instruments allowed the participants to design within an environment based on rules and generative descriptions, amplifying their design understanding and their own learning. The students connected their knowledge with their ambition to create their own design proposals.

The synthesis of all individual projects removed the students from individual ownership of their designs, but allowed them to reflect on both their own and their colleagues' designs as a complete cluster of contributions (Kvan, 2004B). This related to earlier research into design studios based on the same principle, in which media were applied outside their normal pre-described purposes, and innovative design methods were deployed by interplaying digital media and design explorations (Schnabel et al., 2004).

With the employment of parametric design methods that allowed students to experience the dependencies and rules of the various individual contributions spatially, as well as the overall common proposals, the design was communicated using digitally controlled manufacturing processes and digital representations.

The studio was phased in such a way that each section built upon the next and became an essential part of the overall design learning and creation. They addressed and expressed certain aspects of the process. A holistic discussion about design, form, function, and development is consequently established - a significant venture not only within the architectural realm, but also in all other dialogues involving spatial representation.

REFERENCES

Ambrose, M. A. (2009). BIM and comprehensive design studio education. In *Proceedings of the 14th International Conference on Computer Aided Architectural Design Research in Asia* (pp. 757-760). Yunlin, Taiwan: CAADRIA.

Bennetts, R. (2008). Reasserting the architect's position in pursuit of sustainability. In S. Roaf & A. Bairstow (Eds.), *The Oxford Conference: A re-evaluation of education in architecture* (pp. 11-16). Southampton, UK: WIT Press.

Biloria, N., Oosterhuis, K., & Aalbers, C. (2005). Design informatics. In *Smart Architecture: Integration of Digital and Building Technologies* (pp. 226–235). Savannah, GA: ACADIA.

Celani, M. G. C. (2008). Teaching programming to architecture students. *Revista Gestão & Tecnologia de Projetos*, *3*(2), 1–23.

Davidson, P. (2006). *The regular complex.* Paper presented at the NSK Wolfram Science Conference, Washington, DC. Retrieved May 10, 2010, from http://www.wolframscience.com/conference/ 2006/presentations/davidson.html

(2004). *Digital Project* ™. Los Angeles, California: Gehry Technologies.

Dokonal, W., & Hirschberg, U. (Eds.). (2003). *Digital design.* Graz, Austria: eCAADe and Graz University of Technology.

Eastman, C. (2004). New methods of architecture and building. In *Fabrication: Examining the Digital Practice of Architecture* (pp. 20–27). Cambridge, Canada: ACADIA & AIA Technology in Architectural Practice Knowledge Community.

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2008). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors*. New Jersey: John Wiley & Sons.

Forrest, R., La Grange, A., & Yip, N.-m. (2002). Neighborhood in a high rise, high density city: Some observations on contemporary Hong Kong. *The Sociological Review*, *50*(2), 215–240. doi:10.1111/1467-954X.00364

Gao, S., & Kvan, T. (2004). An analysis of problem framing in multiple settings. In Gero, J. (Ed.), *Design computing and cognition* (pp. 117–134). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Karakiewicz, J. (2004). City as a megastructure. In Jenks, M., & Dempsey, N. (Eds.), *Future forms for sustainable cities* (pp. 137–151). Oxford, UK: Architectural Press.

Kvan, T. (2004A). Collaborative design: What is it? *Automation in Construction*, *9*(4), 409–415. doi:10.1016/S0926-5805(99)00025-4

Kvan, T. (2004B). Reasons to stop teaching CAAD. In Chiu, M.-L. (Ed.), *Digital design education* (pp. 66–81). Taipei, Taiwan: Garden City Publishing.

Maver, T. W. (1995). CAAD's seven deadly sins. In M. Tan & R. The (Eds.), *The Global Design Studio, Proceedings CAAD Futures* (pp. 21-22). Singapore: Centre for Advanced Studies in Architecture, National University of Singapore.

Monedero, J. (2000). Parametric design: A review and some experiences. *Automation in Construction*, *9*(4), 369–377. doi:10.1016/S0926-5805(99)00020-5

Oblinger, D. G., & Oblinger, J. L. (Eds.). (2005). *Educating the Net generation*. Washington, DC: Educause.

Oxman, R., & Oxman, R. (Eds.). (2010). Special issue: The new structuralism: Design, engineering and architectural technologies/ *Architectural Design, 80*(4).

Picon, A. (1997). Les annales de la recherche urbaine. *Le Temps du Cyborg dans la Ville Territoire, 77*, 72-77.

Picon, A. (2010). *Digital culture in architecture: An introduction for the design professions*. Basel, Switzerland: Birkhäuser.

Riese, M., & Simmons, M. (2004). The glass office - SCL office and showroom in Brisbane, Australia. In *Fabrication: Examining the digital practice of architecture* (pp. 28–33). Cambridge, Canada: ACADIA & AIA Technology in Architectural Practice Knowledge Community.

Schnabel, M. A., & Bowller, N. (Eds.). (2007). *Disparallel spaces*. Sydney, Australia: The University of Sydney.

Schnabel, M. A., Kvan, T., Kuan, S. K. S. & Li, W. (2004). 3D crossover: Exploring - Objets digitalise. *International Journal of Architectural Computing – IJAC, 2*(4), 475-490.

ADDITIONAL READING

Derix, C. (2009). In-Between Architecture Computation. *International Journal of Architectural Computing – IJAC, 7*(4), 565-586.

Iwamoto, I. (2009). *Digital Fabrications: Architectural and Material Techniques*. New York: Princeton Architectural Press.

Jones, C., Ramanau, R., Cross, S., & Healing, G. (2010). Net generation or Digital Natives: Is there a distinct new generation entering university? *Computers & Education*, *54*(3), 722–732. doi:10.1016/j.compedu.2009.09.022

Karakiewicz, J., Shelton, B., & Kvan, T. (2010). *From Vertical to Volumetric: prototypical Hong Kong*. Routledge.

Rahim, A., & Jamelle, H. (Eds.). (2007). *Elegance-Architectural Design AD. 77(1)*. London: John Wiley & Sons Ltd.

Schumacher, P. (2010). *The Autopoiesis of Architecture*. London: John Wiley & Sons Ltd.

Sterk, T., & Loveridge, R. (Eds.). (2009). *reForm() Proceedings of the 29th annual conference of the Association for Computer Aided Design in Architecture*. Chicago, Illinois: ACADIA

Tidafi, T., & Dorta, T. (Eds.). (2009). *Joining Languages, Cultures and Visions/Joindre Langages, Cultures et Visions - CAADFutures 2009*. Montréal, Canada: Les Presses de l'Université de Montréal.

Wang, X., & Schnabel, M. A. (Eds.). (2009). *Mixed Reality Applications in Architecture, Design, and Construction*. Amsterdam, Netherlands: Springer-Verlag. doi:10.1007/978-1-4020-9088-2

KEY TERMS AND DEFINITIONS

Architectural Computing: Architecture that is aided or generated by computational means.

Computational Architecture: Architecture and its design that arrived from or in collaboration with computational means, instruments or aids.

Design Education: The pedagogical approach to teach and learn to design.

Design Learning: Learning of how to design with the aim to become a good designer.

Design Process: The elements that contribute to the making of a design.

Parametric Designing: Designing using a parametric methodology that employs parameters, rules, and systems.

Parametric Design Studio: Design Studio that employs parametric designing as core method of enquiry.

This work was previously published in Computational Design Methods and Technologies: Applications in CAD, CAM and CAE Education, edited by Ning Gu and Xiangyu Wang, pp. 56-70, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 14
Service Design:
New Methods for Innovating Digital User Experiences for Leisure

Satu Miettinen
Savonia University of Applied Sciences, Finland

ABSTRACT

Service design is establishing itself as a method for developing services and service business. Service needs, new ideas and ways to utilise technology are encountered when the customer and the end user participate in the design process. This chapter focuses on service design methods and the process of how service design can help in innovating customer-orientated service concepts for e-tourism. Service design connects the areas of cultural, social and human interaction. Use of design methods acts as a link between the different views in the service design process. Service design is an emerging field where the terminology and methods are still developing. Mager (2009) has pointed out that the need for service design is evident, as economic development has changed dramatically during the last four decades from manufacturing to provision of information and services. Service design looks at service development from the designer's point of view. Design thinking has the ability to create concepts, solutions and future service experiences for users.

INTRODUCTION TO SERVICE DESIGN THROUGH A CASE STUDY

The tourism industry is a complex area where service design provides a new perspective to service development. In the tourism context service design can be related to at least two approaches: a) Service design as a tool for improving a customer experience, innovating new service opportunities and ROI (return on investment), and b) Service design as a tool for innovating new sustainable service systems and well-being. Stickdorn (2009) writes about tourism as a service-intensive industry depending on the customers' service experiences and their consequent assessments of its quality. There are a number of factors that make tourism service challenging: the importance of service quality in tourism, the high proportion of SME's

DOI: 10.4018/978-1-4666-1945-6.ch014

in the tourism sector, the fragmented constituent parts of tourism product bundles, the significance of expectations and overall customer satisfaction. The design of services has become an increasingly important issue. Service design needs to respond to a new service-dominant logic where companies need to make more attractive value propositions than their competitors. For tourism service providers, superior value propositions rely on the consumers' experiences.

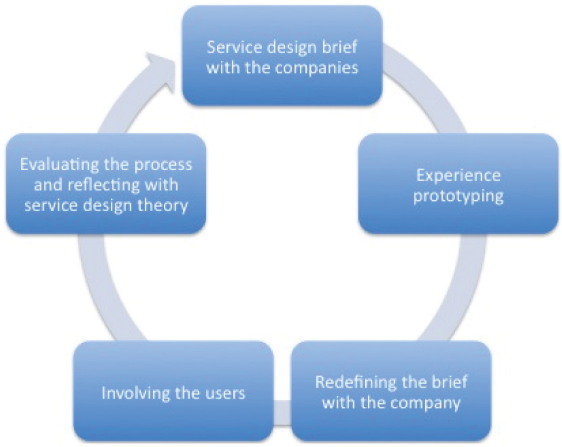
Cipolla (2009) has distinguished a *relational service model,* where the "clients" and "providers" are interwoven. This means that it is difficult to think about *service scripts* or guided (standardised) service performances. As a case she presents an example called "the Living Room Restaurant". This is a service open to everyone, i.e. total strangers are received in a family living room. The "host" family invests their "trust" in others, but there is also a certain level of trust required from "guests": they are entering someone else's place, and it is not known if this unusual service could be a trap. There are mechanisms to relieve this tension: indications about the service trustworthiness can be passed on by word of mouth, for example. The case Living Room Restaurant is a paradigmatic example where we can observe the highest level of trust-making. Relational services propose the achievement of well-being based on interpersonal encounters: an approach that focuses more on "actions" or "relations" than on "things", which leads to sustainable practises. Services that promote ways of living based on sharing and collaboration reinforce the transition towards sustainability: they regenerate the local social fabric and promote the creation of new common goods.

Iteration and co-creation are processes that can also be used to develop services in the tourism area. An iterative design process is based on a cyclic process of prototyping, testing, analysing, and refining work in progress. This applies well in a service design process where prototyping tools are in active use. Innovating opportunities for new co-creation processes between the client and the user is part of the service designer's everyday working life. Prahalad and Ramaswamy (2004) discuss co-creation experiences as new ways to create value. The focus on value in business transactions has shifted to experiences, and experiences are increasingly created through services. Consumers are co-creating value with the firm. Co-creation allows the customer to co-construct the service experience to suit her context, and the service design process offers methods to enable this. Value creation and interaction processes are thus central to service design. Thackara (2008) describes a process where designers transform public services and work with local communities in a co-creation process. Designers have co-developed and prototyped the ideas with the community, and later these ideas have been tested and launched.

The tourism industry brings new challenges to the service design process. Service products are offered by number of service producers as Stickdorn (2009) notes and there may be a lack of communication between the service providers. This means that iteration rounds will be repeated to include all the stakeholders and special attention has to be paid to identifying both the right service channels and all the users of service channels to include them in the co-creation process.

This chapter looks at one specific tourism development case. The case was part of the "Experiencing Well-being – Developing New User Interfaces and Service Platforms for Leisure" project, which was funded by the Finnish Funding Agency for Innovation and Technology, TEKES. The case worked with the development of a future digital service for leisure. This chapter studies how the service design process was used in this case. The aim of the chapter is to learn about the service design process and methods through a practical case study. This was divided into five parts: a) brief, b) experience prototyping, c) redefining the brief, d) involving the users and e) evaluation. This formed an iterative service design process.

Figure 1. Iterative service design process represented in this chapter



BACKGROUND OF SERVICE DESIGN AND TOURISM

Service design has developed as an academic field since the 1990's when the University of Applied Sciences in Cologne first introduced service design as an academic inquiry. Service design has established itself as a part of holistic and innovative design education. It aims to ensure that service interfaces are useful, usable and desirable from the client's point of view and effective, efficient, and distinctive from the supplier's point of view. Service designers visualise, formulate, and choreograph solutions to problems that do not necessarily exist today. They observe and interpret requirements and behavioural patterns, and they transform them into possible future services. (Mager 2009)

Morelli (2002) has defined the service product as a result of interaction between different actors and technological elements during the use phase. He proposes that the design activity should emphasise elements of convergence between several social and technological factors: a) the social, technological, and cultural frames of actors participating in the development of the system, and b) the technological knowledge embedded in the artifacts used for the service. Service design is producing new definitions as well as methods that can be used in tourism service development.

A thorough understanding of the tourism context and the tourist experience can benefit the tourism service design processes as well as the designers who are working with product development, service design and research in the tourism context. The service design process can work as a framework that enables a designer to work in tourism service production. In a tourism experience social interaction between the local producers and tourists play a role. (Miettinen 2007)

Desmet and Hekkert (2007) state that user experience is shaped by the characteristics of the user (personality, skills, background, cultural values and motives) as well as those of the product (shape, texture, colour and behaviour). Physical actions and perceptual and cognitive processes (perceiving, exploring, using, remembering, comparing and understanding) will contribute to the experience. It is further influenced by the context of interaction (physical, social, and economic). Saffer (2007) points out that service design, like systems design, focuses on context. People use products in environments of structured processes. Service design is the design of the whole system of use: the system is the service. Tourism is one of the areas where services are both designed and provided (Miettinen 2007).

Miettinen (2007) has studied the service architecture of a creative tourism experience. Creative tourism shifts the focus from the artifact to the production process of that artifact. The creative tourism experience is a multi-sensuous experience where the feel of materials, smells, the relationship with local craftspeople or other creative producers and their working context becomes an important part of the experience. The creative tourism experience is not only the cultural workshop encounter with local craftspeople but also an experience that commits the tourists in a deeper understanding of the local context. Social contact with local people is constructed through creative experience.

The service architecture intends to describe a user's journey through the various touchpoints (objects, spaces, systems, and people) of the service. People work as significant touchpoints in the creative tourism experience. They deliver the service of teaching crafts skills and several other learning or cultural experiences. They are also the touchpoints that deliver information about the local way of living. This indicates that a robust co-production of service is taking place. The service architecture also shows how the creative tourism experience is divided into pre-, ongoing and post-travel experience (Wang 2000). In this process one user can also pass his or her expectations onto a forthcoming user and create expectations and anticipations. Interaction between different touchpoints is quite linear: one point follows another. This is the nature of tourism and depends on the travel itinerary. The creative tourism workshop experience however is more diverse. There are layered delivery points, and several processes take place at the same time: working with creative craft processes, visiting craftspeople, and learning from the social context. Interaction between different touchpoints is mainly effected through the action of the participants. The participants are telling narratives, inviting for a visit, teaching and sharing skills. These interactions are corporeal and embodied as well as visual and auditory. (Miettinen 2007)

The methodology used for service design research is continually developing as well as the access to analysed information from the tourism development cases where service design methods were used. This knowledge helps to develop both new tourism services and service design theory related to the area.

DEVELOPING A DIGITAL SERVICE FOR LEISURE

This section focuses on a case study on digital service development for a mobile e-tourism service for Hudle oy (http://www.hudle.fi) and local dance festival Kuopio Dance Festival (http://www.kuopiodancefestival.fi). The aim of the development case was to create a new user interface for the needs of the local tourism and leisure industry. Service design methods helped in innovating a new user-driven leisure, tourism and well-being service utilising technology.

Figure 2. The architecture of the creative tourism experience

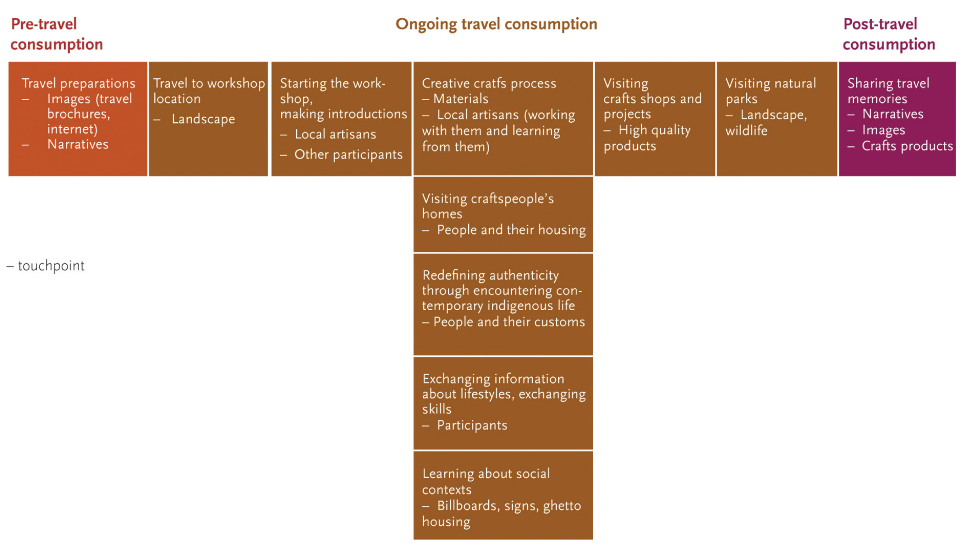


Table 1. A literature review concerning the new research agenda within the service design field

Issue	References	Main contribution
Service design methodology	(Miettinen and Koivisto 2009)	The book presents the key service design terms and methods. It is divided into three areas: an emerging field, service design practise, and case studies. The book discusses the following: a) Design thinking and innovative methods work as tools for co-creating services and desirable value propositions, b) Service design is a tool for designing a more sustainable society, and c) Interaction design offers us insight into creating more user-oriented services.
	(Moritz 2005)	Moritz presents the background on the development of service design, including changes in the economy and in design: how service design can be made accessible in a practical way. It includes service design tools that can be applied in practise.
Service design and improving public services	(Parker and Heapy, 2006)	Parker and Heapy define the service design approach for organisations as a strategy to shape the service organisations around the experiences and interactions of their users. Service designers focus on how people experience services, in order to understand how service organisations can create better relationships with their users and customers. They argue that engaging people in co-production needs to happen at the point of delivery through conversation and dialogue. Learning to understand and map how people experience the point of delivery, the interface with people and their lives, is essential to creating conditions for co-production. Understanding the user is thus the starting point for designing services, and understanding the user experience forges a strong link between the customers and the service design. Parker and Heapy further introduce an interesting goal for service design: to find ways of enabling professionals and laypeople to work together, creating spaces for simultaneous empowerment. Service innovators have focused on creating spaces where professional and personal autonomy can grow simultaneously.
	(Thomas 2008)	Design is key to the challenges of public service transformation. The articles in this pamphlet present a powerful case for design and how it can help achieve better designed public services while fulfilling what can be competing objectives of personalisation and value for money. The contributors to this pamphlet range from social innovators and designers to policy-makers and leaders in government, health and social care. In their articles, they set out the range of projects and innovations where design is already playing a role.
Service design and technology	(Kimbell and Seidel 2008)	The first part of the book focuses on the designing that happens when practitioners within science- and technology-based enterprises work with service designers on projects to design (or re-design) their services. The second area of focus is on how a multi-disciplinary community of practitioners, designers, and academic researchers evolved their understanding of the process and language involved in designing for services.

The service product was developed during an iteration process with stakeholders: Kuopio Tourist Service Ltd (http://www.kuopioinfo.fi/), Rauhanlahti Spa (http://www.rauhalahti.fi), Kuopio Dance Festival and Hudle oy. The first round of negotiations and the design brief was formed in co-operation with these four stakeholders. The first brief focused on service design: a) firstly as a tool to commit the stakeholders in a common service development process, b) secondly as a method to visualise the service where several stakeholders are producing the service together and c) thirdly to prototype the service product. Service design was used as a tool in innovating a new use for existing technology and applying it in the development of a new leisure service.

Service development started with scenario work. The goal of the scenario work was to visualise the service concept with the stakeholders. Scenario work made it possible to focus the design work around certain themes connected with leisure. These themes were drawn out from the analyses of the previous work. Service scenarios were visualised by different methods: using narratives and pictures. The starting point for the project was background research on the area of tourism and leisure as well as the local leisure cluster to benchmark the service selection in the market and development needs of the local tourism cluster. The results of the background research provided the information needed for preparing the context mapping (Sleeswijk Visser, Stappers, van der Lugt & Sanders 2005, Stappers & Sanders 2003) workshop with the stakeholders and users. The aim of the context mapping workshop was to reveal conscious and sub-conscious information about experiences, hopes and expectations related to selected themes. It was easier to discuss and negotiate the service design brief with the stakeholders based on visual material on service scenarios.

The next step of the service development process was the experience prototyping of the service scenario. The aim of the experience prototyping was to test the feasibility of the chosen service scenario: logistics, customer experience and financial influence of the service product in a cheap and quick way. Experience prototyping of the service product also provides much information on the context related to the service experience as well as the architecture of the travel experience: pre-, ongoing and post-experience. This offers a fast and competitive way to realise new customer-orientated service products. (Buchenau and Fulton Suri 2000, Miettinen 2007)

The results of the experience prototyping were analysed with the stakeholders. The video was experienced as a very good prototyping method and it visualised the service concept easily for all the stakeholders. It seemed difficult to commit many service providers in the development process at the same time. Involving users of the various service providers was also challenging for the same reason. We moreover wanted to focus the challenge more in the direction of innovating the technology-based future service, as the prototyping had focused more on existing technology. The group took the decision to carry on the development work with a more focused user group. The service scenario was also focused more on the future.

The development process continued with designing a concept for a future digital service for

Figure 3. The image is taken from the experience prototyping video of the new service product. It depicts a welcome sign for the tourists entering the Kuopio region. It explains to the guests how to join the Kuopio channel using their mobile phones and receive information about different services. The experience prototyping was carried out by video: arriving in Kuopio by car, seeing the sign, accessing the Kuopio channel, using services and special offers of different service providers (restaurant, theatre) and signing out from the channel. (Photograph by Olli Happonen)

leisure. This concept was discussed with Hudle and Kuopio Dance Festival, who agreed on prototyping the concept with the festival visitors. The concept was transformed into an animation. The visitors to the dance festival had the opportunity to see the animation on how the new future service would work and give their feedback. This feedback will guide the future service development.

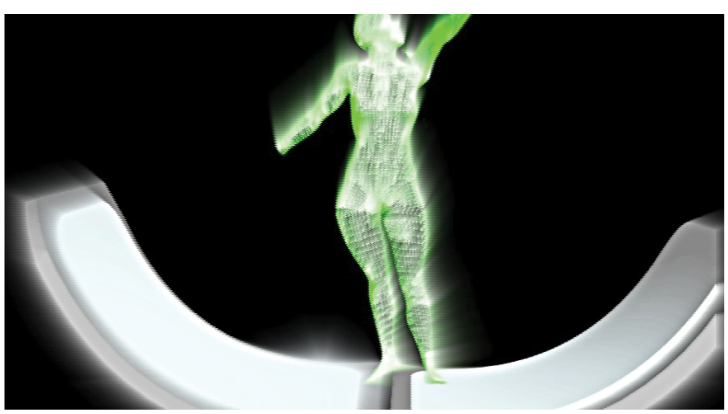
A designer can use a variety of tools to involve the user in the product development process and to acquire user knowledge. Research on user experience offers new kinds of approaches to developing tourism services. The aim of the research project was to create new user interfaces and environments for the needs of the local tourism and leisure industry, and increased knowledge of the user experience and the user are the key elements in designing new services. Existing theory on service design methods and processes helped during the iteration process, finding the right tools and methods. Documentation of the process and the methods is also one way to produce new information about an emerging field.

Experiences from the case study taught us several things. Service design is a tool to commit the stakeholders in the iteration process. It is a process that helps the stakeholders communicate their needs and wishes for the services and better understand the end user's needs. Service design methods can help in concretising abstract service processes with visualising techniques. In this case the video production helped the stakeholders visualise the service in practise: how the end users are able to use the technology and how the stakeholders can financially benefit from the use of technology in their services. Prototyping helped in testing the feasibility of the service concept and developing it further. Involving the users in the service development process transfers knowledge of the user experience in the service and guides the whole development process forward. The end users communicate their experience of the service: if it was credible, enjoyable, easy to use, and so on. This feedback is valuable information that will be fed back into the iteration process.

SERVICE DESIGN IN FUTURE

Service design methods are developing further with increasing research experience and projects as well as with consultancies specialising in service design. In the area of tourism not only technological concepts and service development benefit

Figure 4. The image is taken from the animation of the new service product and depicts a hologram phone: a phone that reflects a hologram image. This phone can produce services for the visitors of the Dance Festival: for example, teach dance steps. This service enables extending the service offering of the Dance Festival both to pre- and post travel experience. (Photograph by Olli Happonen)



from service design but also traditional person-to-person modalities. For example the use of drama methods and enacting the service situations will help the staff to develop their person-to-person service processes, communicate better with the guests and innovate new service opportunities. In future, service design will be a more and more important means in developing tourism services both for committing the stakeholders in the development process and engaging the users/guests in the service development process.

Mager (2009) notes that in these changing markets improvement and innovation of skills are of crucial importance to service providers in order to keep customers and in order to win new ones. Companies need to develop more sensitive antennas for tuning into what today's consumers expect and value, as well as learning what it takes to reach and retain them. Service design is in many cases dealing with human behaviour: behaviour of employees, behaviour of customers and co-customers. Due to co-production customer behaviour is of greatest relevance to the success or failure of the service. Therefore research on the way design can influence behaviour is of major interest within the service design field. The social and public service sector will especially need to focus more on the opportunities of behavioural change through the use of design and service design interventions.

One of the main areas of service design development in future is the public service sector. Tim Brown (2008) from IDEO discusses design thinking and reminds public servants to ask the obvious: What is it like to be admitted into a hospital, call the police or collect the dole? These questions are an excellent start for unlocking innovation by using design tools such as observation and storytelling. Andrea Siodmok (2008) from the Design Council, UK, writes about developing public services with service design. There is a growing pressure on the public sector to deliver "more for less". This makes innovation an imperative, where new ideas will need to be effectively developed, prototyped, tested and implemented. More flexible user-centred services will place greater pressures on many public sector organisations. The good news is that there are many examples of successful design and innovation in the public sector.

It is only natural that designers will contribute to positive societal development more and more using their design thinking and service design tools and methods to create more user-orientated service systems. This will create pressure on design education, and several new design disciplines such as service design education will develop. This will bring new actors into the workforce creating new solutions.

CONCLUSION

This development case was part of a larger research project on service design methods and the main focus was on developing and testing methodology. As such the case study was very fruitful. This development case was also based on earlier knowledge of service design methods and on how the tourism experience is constructed. The main outcome of the case and present research work is that the service design methods are under significant development work. There are different tools and methods, evidenced by many emerging websites and blogs. More systematic academic definitions and research work is needed. The future looks bright and the development work continues both at the academic and the practical level with companies. The most active side of service design has been the fact that the area seems to combine the very active hands-on development with business and methodological enquiry.

Our co-operating partners and stakeholders were enthusiastic about service design and its possible benefits to a company. There are already success stories available as Mager (2009) describes. Virgin Atlantic, for one, has its own department for service design. The Volkswagen research department is integrating a service design approach into their research. McDonald's has set up a customer experience innovation centre using the service design approach as a major facilitator for innovation. Numerous service companies, including telecommunication providers, insurance companies, banks, hospitals, transportation and hospitality industries are integrating service design on an organisational or on a project basis. We also need these success stories at a regional level and with our SME companies who are our partners on a daily basis. The tools and methods need to work also using very small resources.

One of the main tasks of the case study and the research work has been focusing on the results of the iteration process. There is still a lack of literature on service design research. The reflective and discursive evaluation between different research findings and theories is still quite weak. It is important to document and ground the research processes well in order to formalise new theoretical knowledge that can be used in further research projects. The research process has shown that the new methods and tools that service design can offer e-tourism can produce user-orientated service products when the stakeholders are committed in the development work. Yet much more development work is needed at the regional level with SME's that need easy to use and low cost development tools for their work with users. This is our next development phase.

REFERENCES

Brown, T. (2008*). Public servants – by design. The SOLACE Foundation Imprint. The Guardian.* Design Council. Ed. Thomas. E. http://www.solace.org.uk/library_documents/SFI%20-%20Innovation%20by%20design%20in%20public%20services.pdf

Buchenau, M., & Fulton Suri, J. (2000), Experience prototyping. In Proceedings of the DIS 2000 seminar. *Communications of the ACM, 2000*. extools.stanford.edu/papers/SuriExperiencePrototyping.pdf. Downloaded 20.8.2007

Cipolla, C. (2009). Relational services and conviviality. In Miettinen, S., & Koivisto, M. (Eds.), *Designing Services with Innovative Methods. Publication series University of Art and Design Helsinki B 93. Kuopio Academy of Design. Taitemia Publication Series 33*. Otava.

Desmet, P., & Hekkert, P. (2007), Framework of Product Experience. *International Journal of Design, 1* (1), 57-66. http://www.ijdesign.org. Downloaded 4.9.2007

Kimbell, L., & Seidel, V. P. (Eds.). (2008), *Designing for Service – Multidisciplinary Perspectives: Proceedings from the Exploratory Project on Designing for Services in Science and Technology-based Enterprises*. Said Business School University of London. http://www.sbs.ox.ac.uk/D4S/essayArchive/D4S_Publication.pdf

Mager, B. (2009). Service Design as an Emerging Field. In Miettinen, S., & Koivisto, M. (Eds.), *Designing Services with Innovative Methods. Publication series University of Art and Design Helsinki B 93. Kuopio Academy of Design. Taitemia Publication Series 33*. Otava.

Miettinen, S. (2007). *Designing the Creative Tourism Experience. A Service Design Process with Namibian Crafts People*. Publication series of University of Art and Design Helsinki A 81. Doctoral Dissertation. Gummerus kirjapaino oy. Jyväskylä.

Miettinen, S., & Koivisto, M. (2009). *Designing Services with Innovative Methods. Publication series University of Art and Design Helsinki B 93. Kuopio Academy of Design. Taitemia Publication Series 33*. Otava.

Morelli, N. (2002). *Designing Product/Service Systems: A methodological exploration. Design Issues. XVIII(3)*. Cambridge: MIT Press.

Moritz, S. (2005). Service Design: Practical Access to an Evolving Field. *Köln International School of Design. University of Applied Sciences Cologne.* http://stefan-moritz.com/Stefan%20Moritz/Service%20Design_files/Practical%20Access%20to%20Service%20Design.pdf (20.4.2009)

Parker, S., & Heapy, J. (2006). The Journey to the Interface. How Public Service Design Can Connect Users to Reform. London: Demos. http://www.demos.co.uk/publications/thejourneytotheinterface (19.7.2009)

Prahalad, C. K. & Ramaswamy, V. (2004). Co-creation experiences: The next practice in value creation. *Journal of interactive marketing, 18* (3). Summer 2004. DOI: 10.1002/dir.20015

Saffer, D. (2007). *Designing for Interaction: Creating Smart Applications and Clever Devices*. Berkeley, CA: New Riders and AIGA Design Press.

Siodmok, A. (2008). A hotbed of creativity. The SOLACE Foundation Imprint. The Guardian. Design Council. Ed. Thomas, E. http://www.solace.org.uk/library_documents/SFI%20-%20Innovation%20by%20design%20in%20public%20services.pdf 19.9.2009

Sleeswijk Visser, F., & Stappers, P. J. van der Lugt, P. & Sanders, E. B.-N. (2005). Contextmapping: experiences from practice. *CoDesign: International Journal of CoCreation in Design and the Arts, 1* (2).Boca Raton, Fl: Taylor and Francis. http://www.maketools.com/pdfs/Contextmapping_SleeswijkVisseretal_05.pdf (16.8.2009)

Stappers, P. J., & Sanders, E. B.-N. (2003). Generative tools for context mapping: tuning the tools. *Third International Conference on Design & Emotion*. Loughborough:Taylor &Francis. http://www.maketools.com/pdfs/GenerativeTools_Stappers_Sanders_03.pdf (16.8.2009)

Stickdorn, M. (2009). Service Design in Tourism. Designing Services with Innovative Methods. In Miettinen, S., & Koivisto, M. (Eds.), *Publication series University of Art and Design Helsinki B 93. Kuopio Academy of Design. Taitemia Publication Series 33*. Otava.

Thackara, J. (2008). Transforming public services. In In Thomas, E. (Ed.), *Innovation by Design in public services*. The Solace Foundation.

Thomas, E. (Ed.). (2008), *Innovation by Design in public services. The SOLACE Foundation Imprint. The Guardian*. Design Council http://www.solace.org.uk/documents/sfi/SFI%20-%20Innovation%20by%20design%20in%20public%20services.pdf (19.7.2009)

Wang, N. (2000). *Tourism and Modernity: A Sociological Analysis. Pergamon. Tourism Social Science Series*. An imprint of Elsevier Science.

ADDITIONAL READING

Holmlid, S., & Evenson, S. (2007), Prototyping and enacting services: Lessons learned from human-centered methods (PDF file), http://www.ida.liu.se/divisions/hcs/ixs/publications/fulltext/2007/HolmlidEvenson_QUIS07.pdf

House of Commons, Public Administration Select Committee. (2008), User *Involvement in Public Services, Sixth Report of Session 2007–08, Report*, together with formal minutes (PDF file) http://www.publications.parliament.uk/pa/cm200708/cmselect/cmpubadm/410/410.pdf

IDEO. a global design consultancy, http://www.ideo.com/

Leadbeater, C., Bartlett, J., & Gallagher, J. (2008), *Making it Personal. Demos* http://www.demos.co.uk/publications/makingitpersonal

live|work: creates innovative services and improves existing ones, http://www.livework.co.uk/

Morelli, N. (2002). Designing product/service systems. A methodological exploration. *Design Issues*, *18*, 3–17. http://www.aod.aau.dk/staff/nmor/papers/DesignIssues.pdf. doi:10.1162/074793602320223253

Morelli, N. (2003). Product-service systems, a perspective shift for designers: A case study: the design of a telecentre. *Design Studies*, *24*(1), 73–99. http://www.aod.aau.dk/staff/nmor/papers/DesignStudies.pdf. doi:10.1016/S0142-694X(02)00029-7

Service Design Research. aims to collectively build an understanding and foster a dialogue on service design, http://www.servicedesignresearch.com/

Service Design Tools. communication methods supporting design processes, http://www.servicedesigntools.org/

The Service Design Network. an international network for and from organisations and institutions dealing with service design in a professional way http://www.service-design-network.org/

thinkpublic: public service and communication design, http://thinkpublic.com

Wenovski: design thinkers' network, http://wenovski.ning.com/

APPENDIX: CASE STUDY

Develop a Service Channel for Well-Being

Study your local tourism sector and the well-being services it offers: spas and various traditional and new age treatments, physiotherapy, medical treatment and other services that support the guests' personal well-being. You should develop a new service product: "a Service Channel for Well-Being". Keep in mind the service consumption can be extended pre-, ongoing and post- travel experience and think about all these parts of the experience. Start the project by designing and visualising a concept for the service channel.

Questions

1. What are the methods to participate with and identify the stakeholders in the project?
2. How can you involve the users in the development work?
3. Make sure to prototype the channel!

Possible Paper Titles/ Essays

1. Service design and tourism: an emerging area
2. *Creating new service opportunities with service design methods*
3. *Co-creating the tourist experience with service design*
4. Services and new business models in e-tourism

This work was previously published in Digital Culture and E-Tourism: Technologies, Applications and Management Approaches, edited by Miltiadis Lytras, Patricia Ordóñez de Pablos, Ernesto Damiani and Lily Diaz, pp. 36-47, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 15
A Mass Customisation Implementation Model for the Total Design Process of the Fashion System

Bernice Pan
Seamsystemic Design Research, UK

ABSTRACT

Global economic development has been increasingly segregateing the design and manufacturing functions of industries both geographically and administratively. In response, fashion companies in advance economies have increasingly operated as brand houses that engage in design and marketing as their sole activities. The total design processes of the contemporary upstream fashion supply chain is therefore investigated and analysed as an integrated fashion system. A new conceptual model of mass customisation aligning the activities and interests of the collective fashion supply chain producers is subsequently developed with its associated implementation strategies. This model takes a consumer-centric approach, and places designers/brand houses as the instrument and channel for mass customisation. The objective is to enable the prospect for small medium fashion enterprises to deploy the vision and principles of mass customisation in a more coordinated, cost effective, and responsive way. The prospective benefits and practical issues of this new model are discussed.

INTRODUCTION

Fashion companies in advance economies have increasingly become brand houses performing the functions of design and marketing exclusively at two polarised ends of the lengthy and fragmented fashion supply chain (FSC). This includes, firstly, functions of total design processes interacting with off-site manufacturers at the FSC up-stream. Secondly, it includes the functions of marketing and sales interacting (directly or indirectly) with end-consumers at the FSC down-stream.

DOI: 10.4018/978-1-4666-1945-6.ch015

This chapter refers to end-consumers collectively as the fashion 'Customers'. It refers to all participants of the FSC design and manufacturing functions collectively as the fashion 'Producers'. This includes most critically, brand-houses/designers companies, textiles suppliers, and garment manufacturers. The "Fashion System" refers to all workings of the global fashion apparel industries.

Mass customisation (MC) has been recognised as an effective strategy in response to the contemporary climate of "the Fashion System" in an age of diverse Customers demand, product proliferation, global competition and information technology. This climate has been further challenged by the global economic downturn and the impact of currency fluctuations on trading supply and demand. However, MC implementation tends only to be deployed in the mass market sector by global corporations with large scales of economy and extensive technological investment capability, whilst more than 90% of the Producers in the Fashion System are small-medium enterprises (SMEs).

This chapter explores the strategic and practical possibilities of mass customisation for the total design processes (TDP) at the FSC upstream. This includes all processes from design conceptualisation (creative direction) and sourcing to design development and pre-production. The set time conditions of the seasonal calendar are established as the key parameter of the FSC. These time conditions are often neglected in studies of design processes, yet they provide a critical and competitive basis to all works of the Fashion System. The set components and definition of the fashion garment is established as the key output of the FSC. The garment components are what all efforts of the Fashion System ultimately work towards and generate revenues from.

A new model for a mass customised FSC is developed to enable a two-way balance between a consumer-centric demand and a design-led supply process. This balance is critical to the optimisation of the contemporary Fashion System. Due to the process-based principles, the new model does not necessarily require large investment or substantial technological intervention for brand-houses or small medium designers companies to implement, but warrants the benefit of mass customisation.

BACKGROUND

The Fashion Design Process: A Calendar of Prescribed Timing for Supply Push

For fashion brand houses and apparel designers' labels companies, the FSC design process generally includes all stages from ideas generation and creative direction to garment prototype development (Secor, 1992). Much published research explores the design and development process with respective diagrams devised to illustrate key sequential stages.

Carr and Pomeroy (1992) classified four phases with linear operative steps in each: 1) Origin of styles, comprising market research, design, concept, and market screening. 2) Sample development phase, including prototype pattern, sample, and range meeting. 3) Business objective refinement phase, encompassing pattern adaptation and testing, while production planning and scheduling is also carried out. 4) Attainment of commercial products. This takes the process from production pattern, grading, markers, production templates, specification, into two strands of feedback from manufacturing and marketplace. And while technological advances and geographical shifts of fashion design and development have evolved since the 1990s, the process stages and functional output of the FSC processes have remained largely unchanged.

Gaskill's (1992) retail product development model splits activities after 1) Trend analysis and concept development, into three parallel areas of 2A) Fabrication selection, 2B) Palette selection and 2C) Fabric design. This is then followed by

3A) Silhouette and style direction on one hand, and 3B) Prototype construction and analysis on the other, before remerging into a single final phase of 4) Line presentation. The internal factors of target market definition and merchandising process are pointed out, as are the external factors of domestic and foreign markets. However, they are not specified in terms of the stages or time table within which those activities take place.

Wickett *et al.*'s (1999) detailed activities guide for apparel product development, personnel and sources are identified in seven phases of 1) Trend analysis, 2) Concept evolvement, 3) Palette selection, 4) Fabric selection, 5) Fabric design, 6) Silhouette and style directions, 7) Line presentation, with an additional category listing influential internal and external intervening factors.

Sinha's (2001) summarised the generic fashion design process in four phases of 1) research and analysis, 2) synthesis and selection, 3) manufacturing, and 4) distribution, and indicated key factors at each phase. Mc Kelvey and Munslow's (2003) design process flow chart is not dissimilar, though it is completed with an additional promotion phase following completion of range or collection. This would typically be a marketing function carried out or directed exclusively by the brand house, baring no relation or responsibility to the manufacturers.

Forza and Vinelli's (2000) research describes sequential activities in the FSC and points out the substantial reliance design and productions has on forecasting, however inaccurate. Information on customers is typically managed completely independent of activities in the FSC. Their production stage chart illustrates the occurrence of monthly activities on a larger scale from yarn to textile to apparel production and distribution, though does not address specific time factor or relationships between the activities and processes. Le Pechoux's (2001) on the other hand, explores the dynamics and processes of fashion industries as a complex system; addressing the non-linear nature of cyclical iteration processes involving external and internal factors.

Senanayake and Little (2001) reviewed various mechanisms for the textiles and apparel industry's new product development (NPD) measures, and referred to the merchandising calendar in the format of a scheduling Gantt chart, with a full list of 40 activities and respective timing in a total of 35 weeks from concept to production for one season's product range.

Pan's (2006) research addressed a full range of parallel functions in the FSC processes with reference to the fashion calendar timeframe and main interacting participants amongst the collective Producers. An Amalgamated Process Map was devised to show a typical 16-18 wks development period from concept to completion of a full range of up to 200/300 show prototypes for brand houses that produce 2 main seasonal collections per year. The 'Producers' selection point' and the 'Customers' choice point' for the critical total design processes of the FSC were identified.

All process models reviewed reveal that the lengthy total design processes carried out collectively by 'Producers' prior to launching products in the market place, involve a series of screening and selection of design options and combinations. These range from ideas, trends and moods to the palettes of fabrics, colours, silhouettes, and fabrication to the final merchandise stories, mix and timing. Judgment calls are made by Producers loosely according to the evaluation of trend forecasting, observation of market happenings, assessment of past sales, and availability of resources. With the exception of resources availability, all factors considered by the Producers in fact contribute towards a high-risk speculation of what Customers may want in months to come. In this sense, Producers act on behalf of Customers to pre-empt decisions on what they should want.

The cyclical pressures to achieve the right decisions mix occur at set calendar times (typically just before the deadline for the pressing commencement of each following process stage) at one polar end of the FSC upstream, some 6-12 months prior to fashion collection launches at the market place. These decisions therefore directly impinge on the marketing and sales activities: affecting the responses from industry buyers and press at international trade shows, and eventually determining the Customer driven retail sales performance at various regional markets across the globe. It is therefore fair to categories the contemporary Fashion System ultimately as one that exercises a strong "Producers" supply-push approach, with "Customers" well shut-out of the loop of its core design functions and processes.

THE MASS CUSTOMISATION TENET

A Dilemma in the Power Struggle of Demand Pull

Mass Customisation (MC) has since the 1980s been advocated as the prospective paradigm that combines the benefit of mass production efficiency and craft production individualization. The aim is to effectively respond to global turbulence in a market of ever-fickle desires of individual customers, ever faster moving trends, and ever finer segment differentiation, indeed a climate of today. Gilmore and Pine II who pioneered the theory of MC defined it as 'markets of one', whilst Davis defined it as 'individual segmentation' (Davis, 1996; Kara & Kaynak, 1997). Brannon (2000) efines in *Fashion Forecasting*, as the medium between mass production's 'one-size-fits-all' and haute-couture's 'one-one-a-kind', where traditional supply chain is inverted by making Customers an integral part of the initial stages of product development.

Kay declared as early as 1993 that it is possible for all aspects of production and delivery systems to be done effectively to meet Customers' individual needs with the following formula: Mass customisation = custom-focused x lean production x continuous variety x short cycle development x flat and empowered management culture (Kay, 1993). Thackara (1998) further defines the relationship between suppliers and customers in '*One-to-one on-line: fashion, the internet, and the role of design*', outlining how mass customisation provides the platform on which the sales transaction is the beginning of an inter-dependent relationship between the seller and buyer as opposed to the end of a marketing process. This approach repositions functionality and quality as merely a starting point. In a contemporary market place where Customers are increasingly spoiled by the wide range of choices of more-or-less perfectly performing products, it is the logical response that service becomes part of product within a holistic concept for organisational design.

However, despite extensive academic and industry research providing conceptual and practical guidelines, benchmarks, and matrixes for MC, propelled by the challenging economic climate and maturing Customers' behaviour that beckons MC (particularly in the notoriously fast-moving Fashion System), the paradigm is still far from a wide-reaching reality some two to three decades later. Tseng and Pillar (2003) believe the missing gap is the capacity and mentality within companies to put the demands and wishes of each single customer at the centre of the value chain. On the other hand, Blecker & Abdelkafit (2006) argues that there is still a lack of common framework concerning the effective implementation strategy to aid MC practice.

Six levels of mass customisation involving Customers at varying degrees in the Producers' processes can be classified as below with respective contemporary fashion case examples. Level one has the most detached relationship between the Producers and Customers and level six has the most engaged. As the level of customisation increases, Customers involvement moves further upstream of the supply chain (Lampel & Mintzberg, 1996; Blecker & Abdekafit, 2006).

1. The 'Basic Customiser' combines 'push-through' forces of strong market understanding with responsive manufacturing systems,

and expects customers to find what they want from a larger variety of product offering. This is typically an approach taken by large scale companies whom have the resources to develop, buy-in, and offer vast product assortments and design lines. Spanish retail giant Zara is a well-know case whose fast reaction towards seasonal fashion trends, effective use of sales information, and efficient manufacturing system stunned the global industries with its phenomenal success. Moreover, its impact forced other large retailers (particularly US ones) to re-examine their own processes.

Aggressive British large retailer Top Shop is another successful 'push-through' example, using effective marketing and buying strategies as well as exclusive design commissions to curate a wide variety of on-trend products and smaller designers' labels. It caters to the tastes and styling of a wide range of younger age-group Customers', yet fosters their loyalty to Top Shop as an umbrella brand and destination.

2. The 'Alteration Customiser' employs a third party between Producer and Customers to carry out the customisation or personalisation of standard products. Examples are commonly seen in up-scale department stores such as Harrods or Selfridges, where in-house tailors or seamstresses would be called in to make minor garment alterations necessary to fit the specific customer making the purchase.

A small number of designers' boutiques or menswear suiting retailers also have such function or have recommended subcontractors or suppliers whom provide such service to their customers. It is a 'reactive' and an expensive way of obtaining customization (for Customers in above case examples) as products are not designed for effective customisation (Tsigkas et. al., 2001).

3. The 'Cosmetic Customiser' deals only with packaging, distribution or point-of-sale changes. For instance, Thomas Pink shirting company is able to offer customers the service of embroidered personal initials on the cuffs of shirts purchased from its standard ready-to-wear range.

Growing high-end fashion internet retailer Net-a-Porter makes the simple gesture of a 'Cosmetic Customer' by providing 4 choices of packaging options of basic, discreet, signature, and wedding wrap during the check-out point on-line Customers' purchases.

4. The 'Transparent Customiser' provides unique products by monitoring customers and without necessarily making its customisation explicit (Anderson-Connell et. al, 2002), as opposed to a user making the request. The iconic internet retailer Amazon of the internet era is a pioneer and leader of such Customisers, where a full purchasing history of Customers are recorded in its database, and repeated customers are recommended a range of different products in similar styles or genre to what they had previously purchased.

Many internet retailers now employs this effective methodology and Koodos is a designers brand discount site which recommends complementing garment separates or accessories in similar price points or styles Customers shop on site.

5. The 'Self-Adaptive Customiser' produces products with features designed into them so that Customers can change aspects of it according to their individual preferences without aid from manufacturers or retailers. Common examples would be seen in a reversible jacket or a pair of cargo pants where leg parts can be removed and worn as shorts.

Elliot Rhodes takes the 'self-adaptive' principle to design and retails a wide range of fashion belts where Customers can select, self-assemble and interchange belt buckles and embellishment from a standard range to accessories for a variety of fashion and functions.

6. The 'Collaborative Customiser' (Pure Customiser) involves Customers into the product design process termed co-design (Fiore et. al, 2000, Anderson-Connell et. al, 2002). This is the most sophisticated level of mass customization, also known as core/pure customization. This approach nurtures a 'Prosumer' (Toffler, 1980; Piller, 2003) relationship between Producers and Consumers, involving Customers within the design development and/or manufacturing process. A Suit That Fits.Com deploys the classic made-to-measure technology to enable customers to select from their standard range of styles, cuts, fabrics, and finishes to assemble and deliver the product of Customers' personal choice.

DEPLOY demi-couture is a pioneering case in the high-end fashion sector, where Consumers' requests or feedback are incorporated into the design and merchandising processes at the FSC upstream. Garments are designed and constructed with multiple functions, styles, and/or specific customer requirements so that Customers can customise each outfit for numerous looks and occasions. Product life cycle is simultaneously increased as the Prosumer relationship is extended after-sale as loyal Customers build-up a much more versatile and creative personalised wardrobe.

Alford (2000) suggests that the MC concept is suitable for low-volume applications with customisation available at a cost premium from the standard product. Okonkwo (2007), however, argues that the rules for the fashion market place have changed and that customisation need not necessarily be applied at a cost premium to the brand-houses. And in fact, Customers are increasingly expecting and not just desiring personalised attention from their goods/service providers in the following lateral areas: 1) Customising of standard products, 2) Customising point of delivery, 3) Customising of retail shopping experience, 4) Producing bespoke goods, 5) Customising of online shopping experience, 6) Allowing Customers to customise the process.

Whilst there is no shortage of concurrent views for key drivers and market suitability of mass customisation, namely: 1) heterogeneous demand, 2) short product life cycles, 3) mature markets and, 4) more conscious Customers (Kotler, 1989; Pine, 1993; Westbrook & Williams, 1993; Hart, 1995; Kotha, 1995; Feitzenger & Lee, 1997; Gilmore & Pine, 1997; Peppers & Rogers, 1997; Bardakci & Whitelock, 2003), mass customisation in its higher-level and purer forms, is yet to be substantially adapted by the Fashion System on a broader basis for a wider consumer base. More importantly, MC confronts the Fashion System with several of its inherent paradoxical attributes, particularly in the high-end fashion sector. It has thus has been met with skepticism and with substantial reluctance for experimentation and ultimate change. The reasons can be summarised through the following dilemmas that constantly thrusts the FSC towards mass customisation in concept, yet consistently drag it away from it in practice:

Perception Hindrance

Whilst varied product ranges and features are now clearly crucial in meeting an increasingly knowledgeable contemporary Customers' demands; whilst individual tailoring is an elitist's luxury too expensive to sustain both for Producers and Customers, and whilst Customers' willingness to spend extra for additional individuality or personalised convenience has become well known (Taylor et. al., 2003) – the attributes of mass customization would offer an attractive response.

Ironically however, even as the fashion Producers strive to gain greater market share and customer base 'en-mass' in a globalised economy, a remarkable proportion of the Fashion System practitioners are unable to reconcile the increased individuality afforded by mass customisation with warranted conformity offered by 'exclusivity' (e.g., luxury goods baring the same look and logos sold in mass quantities globally). This is driven by a fear of loosing their brand superiority and authority through enabling Customers to specify or personalise their own branded products (Okonkwo, 2007).

Resource Hindrance

Whilst the fashion life cycles commanded by seasonal changes are naturally short, fierce global competition has further propelled ever changing trends and shifting niches, and thus ever shorter product life cycles and un-salvable product features through 'planned obsolescence' (Easey, 1995; Jones, 2002). In full view is the front end momentary glory of the latests, the newests and the musts-haves, fully aided by the contemporary fashion's all powerful counterpart – the media.

This phenomenon would have logically beckoned the practice of mass cutomisations, as it has in many other industries. However, un-like most other industries, fashion brand houses and designs companies in advance economies by-and-large have very low entry barrier and require relatively low technological investment, particularly since shedding almost all of their manufacturing activities some two decades ago. This factor coupled with the apparent necessity to allocate hefty marketing budgets in order to succeed, means both financial and human resources for substantial back-end development, innovation or upgrade commands low or no priority.

Operation Hindrance

In an inherently labour-intensive industry, where each production run of a product style was ultimately never long enough for even a mass retailer in its mass production hay-day to fully automate garment manufacturing (partially due to the nature of handling wide varieties of soft fabrics); where operational skills of individual workforce is heavily relied upon; where competition fierce, entry barrier low, and the need to distinguish one product from a vast array of others paramount; a shift to mass customized production would have seemed both natural and slight.

However, the ever compressed design-development time for the 'new season' revolving chains of cyclic deadlines up from Designers, Textiles Manufacturers, and Garment Manufacturers and down to trade and consumer press, marketing and sales, means that there is never the window or aspiration to reflect, review, or restructure for substantial or collective changes. In order to implement mass customisation effectively, a period of time would be required to coordinate supporting technology and methodology, and devise, update, improve or revise the relevant processes.

Nevertheless, early adapters of MC in the Fashion System, such as Levi's made-to-measure jeans and Nike ID shoes, have maintained their leading positions and proven competitive advantage in the global market place a decade on. They have invested heavily in enabling technology (and marketing) in relation to Customers and Producers. However, studies on these cases have also concluded that success will depend on customers' satisfaction level during early pre-manufacturing steps, and suggested that FSC may need specialised MC processes (Lee & Chen, 1999).

Alford's research concluded that it is possible to transform the management of the customised supply chain by increasing involvement of suppliers and use of product data management systems to effectively combine resources and expertise in geographically dispersed areas. It will also help cope with escalating complexities in managing the varieties and volume of data that mass customisation introduces into the supply chain (Alford et. al. 2000). However, it was identified that optimisation requires an effective and coordinated approach to inform and support relationships and management decisions in the design and production process.

Tseng and Piller (2003) further suggests that building a value chain that places individual Customers demands and wishes at the centre implies much more than investing in technologies. By presenting Customers' co-design prospects, process could be shortened and streamlined to focus on proving real value to Customers, thus increasing competitive advantage to Producers. All authors of MC research agree that industries will greatly benefit from knowledge and utilisation of Customers' information (i.e. real demand), and that much work is to be done throughout the supply chains in order to realise the true gains of MC. The how and where of this balance between Producers' supply push and Customers' demand pull in a new MC process specific to the Fashion System is therefore the central objective of this research.

THE PARAMETER AND OBJECT OF THE FASHION SYSTEM

It is clear that the common objective of the FSC Producers is to produce clothes from conceptualisation to consumption speedily, profitably, and to the Customers' purchasing choice. This research therefore disassembles the current FSC design processes into the following factors with subsequent findings:

The 3 'Constants' of the Fashion Parameter

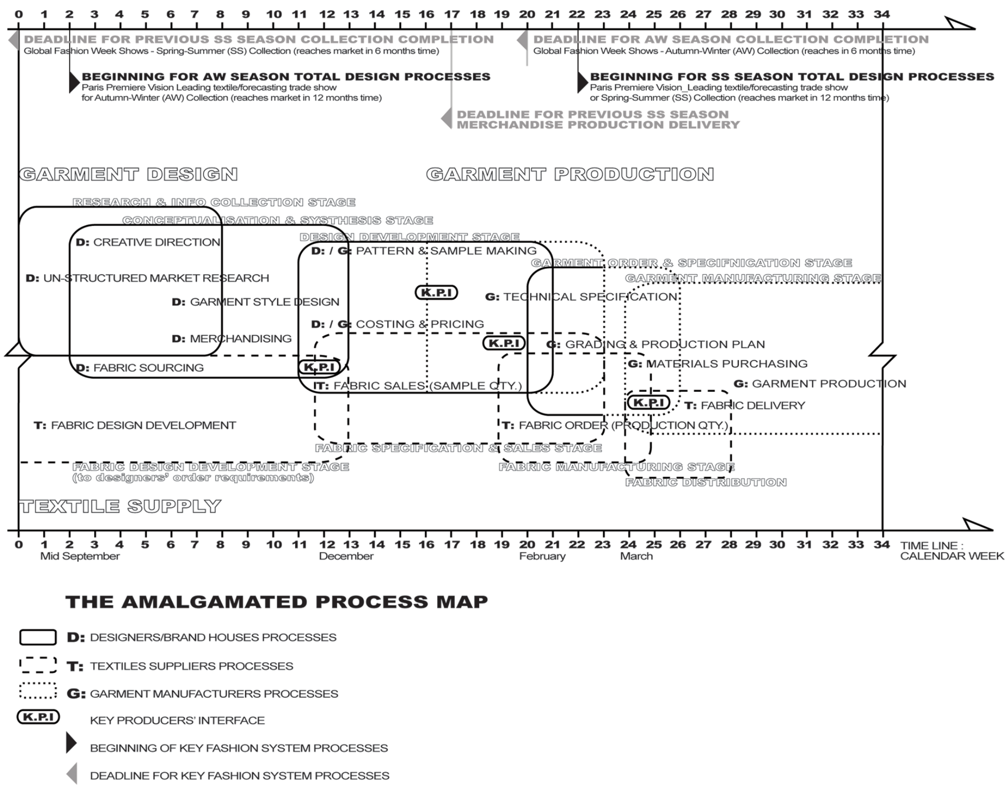
The three key conditions of the Fashion System to which all Producers are subject to, are identified as: 1) the time line, 2) the communication plane, and 3) the decisions points of the FSC total design process. This research therefore considers the three factors as 'constants' of the process model.

Time Line

The parameter of time line is marked into two types of effective periods: 'processing time zone' and 'interacting time zone'. These definitions set the first parameter for the model, and two terminologies refer to the activities segments that must be carried out by the three key Producers of Designers/Brand houses, textiles suppliers, and garment manufacturers. The reference of calendar weeks in the fashion season is common to all parties. Processing time zone refers to the periods where the back-end supporting and preparatory works are carried out in order for the essential communication to take place between the three Producers parties. This typically occupies in a longer time span of calendar weeks. Interacting time zone occupies a shorter time span and refers to the periods where the actual contact and communication is made at the front-end between FSC parties. Zipkin (2001) refers to the activities of this 'interacting time zone' as the 'elicitation process', and categorised the key information as identification, selection from menu of alternatives, technical measurements, and reaction to prototypes.

The key steps of the FSC total design process are shown in the Amalgamated Process Map (Figure 1). All activities for embarking on a new season collection or range design begin officially with fabric selection at or around the major textiles and trend forecasting tradeshows in mid September for the following year's fall-winter fashion collection. The most notable leader for all mid to high-end fashion sector Producers is Premier Vision trend forecasting and textiles tradeshow in Paris. This calendar time to which this tradeshow takes place is thus denoted as week 1 of the Amalgamated Process Map. The same design process is repeated at a maximum

Figure 1. The amalgamated process map



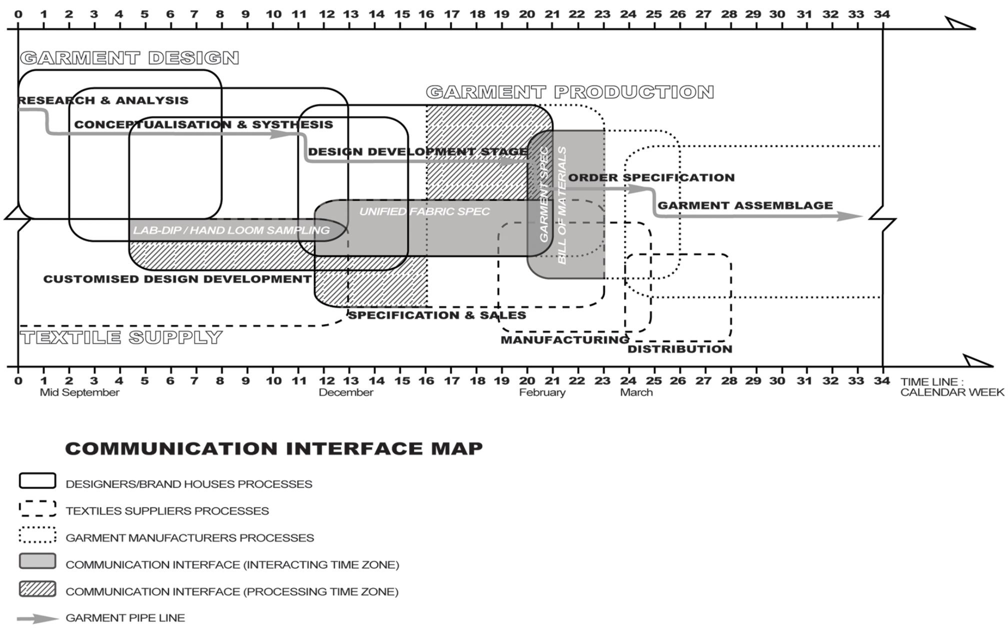
of once every 6 months (each 'fashion season'), and starts again in February for spring summer collections for its following year.

Communication Interface

The design and control of design development process is often ad-hoc based on individual managers' experience without sufficient communication structure (Tsigkas *et. al.*, 2001). Field research analysis showed that smooth and accurate communication between the three key Producers parties carried out by their respective front-end teams in the 'interacting time zone', depends heavily on the correct design information relayed to, and prepared by, the back-end support team in the 'processing time zone'. It also shows that the nature, format, and to some extent, the content being communicated, are in fact the same season after season. These being: accurate and consistent samples and specification packages, and bill-of-materials for both fabrics articles and garments styles.

It is clear that the reduction of information duplication and repeated procedures can improve efficiency and effectiveness thus speeds up processing time for all parties in the FSC. Furthermore, sharing of correct information accelerates problem solving cycles and enables the anticipation of potential issues and obstacles during the design development processes (Muffato & Roveda, 1999). The communication interface (Figure 2) between specific operating teams of the three parties is therefore established as the second parameter for the model. The communication interface is defined by the unification of fabric and garment specifications in terms of the channels, format, and information content for

Figure 2. Communication interface map



interacting FSC parties, with reference to timing requirement and duration allowances set in the fashion system time line.

Decision Integration

Key decisions for both the Producers' selection point and the Customers' choice point (PSP and CCP in Decision Point Map) are converged to become an 'Integrated decision point' (Figure 3). These are the points in time (between weeks 8 to 18) when the most critical decisions for the fashion output of garments are made, from colour and fabric palettes to silhouettes and merchandise mix. They are therefore the specific points where information on Customers' preferences and requirements should be incorporated into Producers' decisions in the relevant FSC process stages.

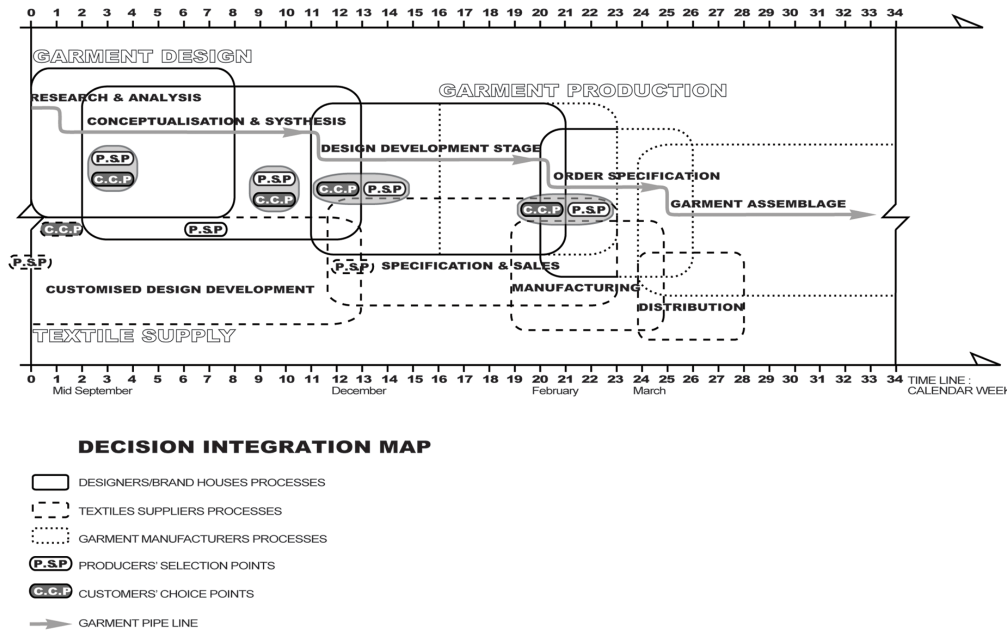
The 3 'Variables' of the Fashion Object

The 3 key components of the Fashion Object, i.e., the garment, to which all 'Producers' output ultimately go towards, are identified as 1) fabric, 2) colour, and 3) silhouette. It is through the incessant changes of these very factors that variety and newness of garments are created season after season to supply Customers. This research therefore considers these factors as 'variables' of the model.

Fabrics

Fabric is undoubtedly one of the first key components of a garment. Fabric selection in the total design process typically entails the pooling of all sample swatches collected from trade shows and individual suppliers' presentations or from Designers' own sourcing and market research. Designers then decide on a palette of desirable fabric appearance, construction, hand-feel, and special features or functions to compile a creative

Figure 3. Decision integration map



direction 'mood board' and/or start silhouette designs with.

In the field research conducted, a range of fabric specifications sent from high-profile, international sample Designers/Brand houses to textiles suppliers were collected and analysed. The specifications showed a wide variety of formats and quality of information. Some were with fragmented hand-written descriptions accompanied by photocopied art work and sample texture of fabric requirement. Others saw formulaic files of large retailers containing lengthy technical lists stretching up to 3 pages, yet with large areas or information irrelevant or not-applicable to most fabrics commonly used, hence barely filled in. Researcher therefore identified information common to all companies and crucial to textiles suppliers' proceedings for fabric production, with references to the following: 1) literature on textiles manufacturing processes, 2) field research on textiles suppliers' order processing and production steps, and 3) key sample textiles suppliers' fabric library cataloguing systems and information referencing methods.

A proposed 'unified fabric specification' with clear information on purchase order, fabric construction and end-use (garment) requirements was subsequently devised in order to assist designers/brand-houses to accurately and efficiently communicate their requirements to their textiles suppliers. Relevant personnel at the textiles manufactures responsible for fabric order processing, as well as industry experts in textiles development were consulted to verify the validity and practicality of the 'unified fabric specification' as a communication interface.

Colours

Colour is a most critical factor in garment design that dictates and reflects the changing seasons and trends in the Fashion System. Whilst Customers may not be fixed on fabric or silhouette (cut) when shopping, most have instinctive and clear ideas on colours, particularly when choosing apparel goods set against their own complexion (Wilson *et. al.* 2001). Field research on consumer's clothes shopping preference showed that 63% surveyed women find colour to be the first thing that catches their attention.

The lengthy series of Producers' 'pre' selection and screening processes for colours take place prior or in parallel to fabric palette selection. They are carried out first by global trend forecasting agencies and fibre, yarn, textile manufacturers, then by Designers/Brand houses, industry buyers, followed by media/press. However it is the final purchasing decision of Customers that determine the success of a colour choice.

From the Designers/Brand houses' point of view, consequences of getting the colour palette wrong could result in costly retention of unsold garment units and loosing market shares to competitors (Wilson et. al. 2001). Instrumental in designer's process in arriving at the final colour choice is the Textiles Suppliers' responsiveness and capability of providing custom physical colour lab-dip match samples or hand-loom swatches as accurately and quickly as possible.

From the textiles suppliers' point of view, the colour specification issue is a complex and costly one and the chances of not getting the colour 'right' in the eyes of the client (Designers/Brand houses) are extremely high. This is due to the following: 1) the subtle differences and subjective preferences in match colour acceptance level, 2) the same chemical colour formula will result differently on different types of yarn, construction, and finishing, and 3) discrepancy in light sources between Textiles Suppliers and their clients'. Even when high-quality textiles suppliers typically provide 3-4 derivatives for one colour match in each lab-dip package, research shows that more than 70% of the colour match samples are approved only after 3 tries of lab-dips (with an average turnaround time of 2-3 week for each 'try'). In addition, Designers/Brand-houses often take up to 2 weeks to reply with approval or request for further lab-dips, due to in-house communication and decision between relevant design or management personnel. It is therefore clear how these reiterative processes lead to an in-efficient 8-10 week period to conclude on fabric development and lengthens the FSC total design process, as marked in the Amalgamated Process Map (Figure 1).

Silhouettes

Silhouette is referred to as the three dimensional cut, overall shape and volume of the design (Mc Kelvy & Munslow, 2003). For the ease of description, Researchers refer to the silhouette factor in the garment design process as a term which comprises the aspects of garment cut; shape; style; construction; proportion, specified through pattern-making and garment specification.

During the period of the fabric and colour selections and decisions, the total creative directions and styling for the seasonal collection/range are conceptualised. The actual designs (drawings) of garment silhouettes (styles) typically begin 2-3 weeks later by Designers. This opens up some 4-6 weeks of garment designing and is carried out concurrently with the continuation of fabric and colour development screening and selection aforementioned. Pattern makers on or off-site then translate drawings with descriptive notation for detailing, finishes and measurement requirements into flat patterns for on or off-site garment sampling.

Design Synthesis

At the end of the parallel processes described in 3.2.1-3.2.3, a set of composite decisions based on three variables above are made on behalf of end Customers by Designers alone in SMEs, or collectively with their in-house textiles and merchandise managers, their company executives in larger fashion corporations. This takes place at the 'collection presentation' process step during weeks 14-18. The collective Producers' aim is that at first sight (of physical retail garment or marketing campaign imagery), the colours, textures, and shapes will attract Customers enough for them to want to buy.

As observed from all sample Designers/Brand houses, a fair percentage of garment silhouette in each company's season collections are repeated designs or consist of similar core shapes from previous collections. This may not always be immediately apparent due to different usages of fabric, colours, and embellishing details. However, further research on the design development processes showed two key evidences for common occurrence of silhouette design repetition.

Firstly, styles in new season collection development are often referred to as a certain previous season style name and/or number with the new addition of a certain detail, both in Designers' verbal descriptions and sketch notations. At times descriptions of such nature replace formal drawings or initial specification altogether. Secondly, certain key pattern blocks are frequently re-used for different garments with minor alteration or addition of lines and details to generate new season silhouettes. This was seen particularly often in the work of Designers/Brand house with in-house pattern makers, where archive patterns were frequently accessed for referencing purposes during the construction of new styles. However, the practice of silhouette repeats and moderate alteration is dependent upon individual Designers'/Brand-houses' preferences and work methods, and this practice is not carried out in a systematic or consistent approach in any of the sample companies observed.

The main reasons that the design of garment silhouettes often repeat from season to season, as widely seen in the market place in general, are as following: A) Long-term over-arching trends on silhouettes last for several years and through the seasons (Jackson, 2001). B) Each designer/brand house develops certain garment cut and fits typical to its products that are recognised by their clientele and thus maintained as classic items carried over the seasons. C) Although there are countless looks and styles to which the Fashion System generates constantly, between the variety of fabrics and colour palettes, there are arguably a finite number of key silhouettes amongst each garment type that would be accepted by a wider clientele. For instance, there is a limit to which the silhouette of a pair of full length trousers can vary, before its potential buyer group is dramatically reduced (irrespective of price points), and sales prospect compromised.

Nevertheless, repetition in garment silhouettes, wholly or partially, is by no means a lesser expression of creative design, for it is the combination of the three variables of fabric, colour, and silhouette that generates desired value and newness for Customers. The example of the famous Burberry checks print used on typical swimwear fabric for a basic bikini silhouette illustrates that high value can be re-created by altering just one variable of a design. In fact, when closely examined, most garment silhouettes consist of some elements where aspects are not created completely from scratch and have not or could not form parts of another garment silhouette.

This dissected view of the garment object highlights the significance of systemic repetition in the contexts of design process. That is, if within in each garment silhouette there is always some portion of it re-produced from parts of another garment previously designed, it takes no great leap towards two possible prospects. Firstly, the generating and communicating of largely the same information of garment pattern and specification can be logically organised for frequent access and cross-referencing. This would reduce the common occurrence of repeat information and/or process duplications as widely observed in the FSC processes, thus allow for a more streamlined and sustainable process saving both time and resources. Secondly, components for repeatedly designing garments with partially shared silhouettes, (e.g. jacket of same bodice cut yet different lapels and sleeves shapes), can be methodically rationalised and utilised – thereby capitalising on the natural existence and characteristics of silhouette repetition as a design variable, without compromising the creative totality of a garment.

CONCLUSION AND DISCUSSION

Proposed New Process Model and Implementation Strategy

The Collective Activity Model (Figure 4) which maps out the total design process emerges as a new prospective model of mass customisation for the FSC. The new model is characterised by the following attributes: firstly, the rationalised and directional flow between Producers and Customers is cross-referenced against the calendar time line of the existing Fashion System. Secondly, Designers/Brand houses are positioned in the critical role of integrating Producers' design-development processes at the back-end and Customers' real choices collated directly from the market retail front end. Thirdly, the Customers' information is incorporated into 4 key points in the FSC upstream process; points 1 and 2 are 'opinion-based' information that aid Producers' final colour, fabric, and silhouette selections. Points 3 and 4 are 'decision-based' information that aid Producers' sizing and quantity order. Fourthly, the moderate changes to the collective FSC upstream processes do not require all participants of the fragmented FSC to radically re-engineer their existing functions and processes. Finally, the model allows for further compression of MC delivery lead-time by systemising aforementioned communication interfaces and design synthesis.

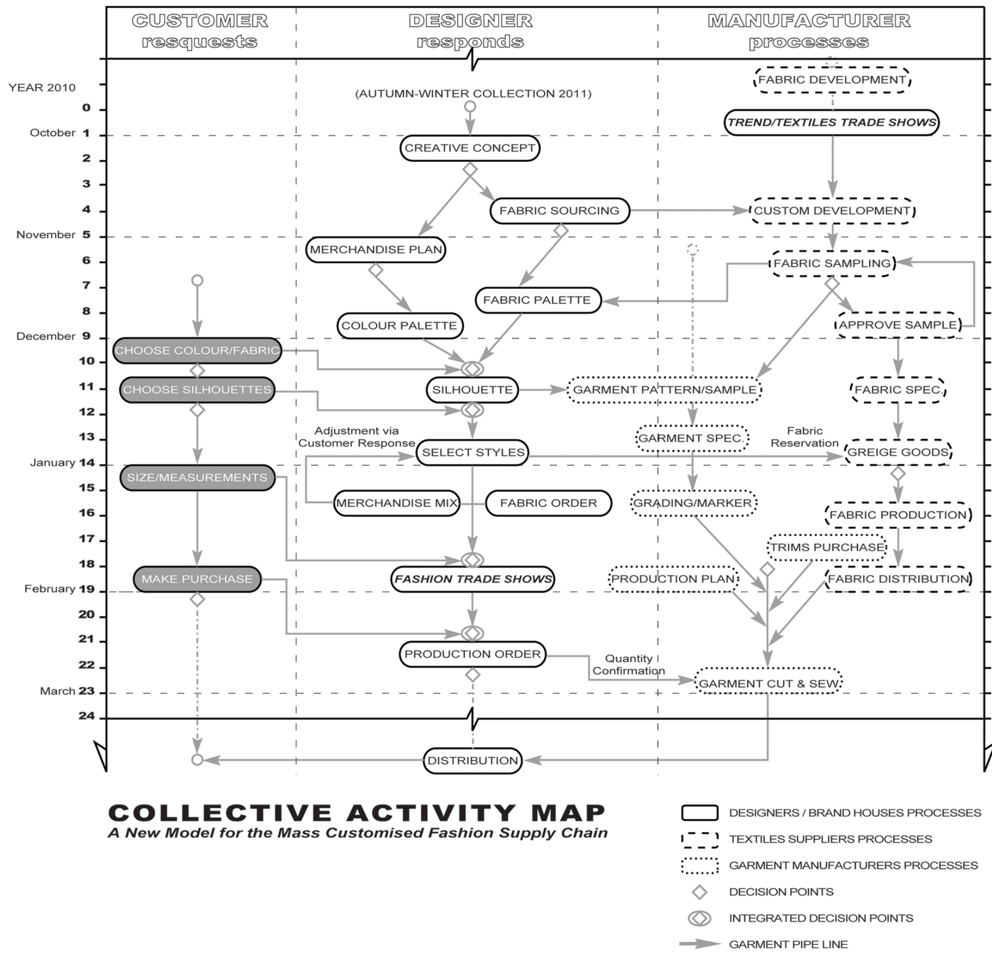
The new model therefore offers the potential for Producers to supply the garment in a more accurate manner, and for Customers to demand the garment in a more personalised fashion. However, it does not require significant organisational re-structure and technological investment, nor presents the high risks that would prevent fashion SMEs from experimenting with this new model for mass customisation whilst constantly under the seasonal time pressures. Three key implementation strategies are discussed in the following sections with reference to Customers and Producers. In this new model Producers are differentiated into Manufacturers, and Designers for clarity of the strategic propositions.

On Consumption: Customer Requests

With a central focus on mass customisation, the field research survey and analysis on sample female shoppers in central London investigated what and how the fashion Customers demand, and what their unfilled demands are. The results cast light upon its reciprocal relationship with supply. From the Customers' point of view, most Producers are unresponsive and insensitive towards their precise needs and wants, despite the enormous investment to attract Customers through Designers/Brand-houses marketing efforts. The survey showed that 64% of respondents wished they could easily access customised garments, and a total of 63% of respondents have clothing alteration needs 'frequently' or 'all the time'.

This unresponsiveness to Customers (from garment style, to colours, fit, availability and price) translates directly into lost sales on a significant scale and in turn results in loss of market share for the Producers. The consequences account for two facts: firstly, in a stiffly competitive, densely populated, yet marginally differentiated fashion system, customers can find easy substitutes for most fashion trends and merchandise. Secondly, due to widespread information technology, customers are fast becoming knowledgeable about fashion trends and products. They are thus increasingly able to declare what they like and dislike, and to even specify what they need and want. This illustrates their full readiness for all levels of mass customisation offering.

Figure 4. Collective activity model: A mass customised FSC



New Model Strategy 1: Customers' information input built-in

Whilst theories of postponement and modularisation for mass customisation have previously separated the production-driven push system and the customer-driven pull system, and identified the 'decoupling point' in the value chain (Blecker & Abdelkafi, 2006) from which standardised and customised products are separated in the pipeline, this research emphasizes the importance of a combination of supply-push and demand-pull approach for the FSC. 'Integration Decision Points' are identified in the total design process with the precise timing and contents necessary. Customers' feedback and general preferences or wish-lists (particularly that of repeat Customers) on product attributes such as fabrics, colours, and silhouettes can be continuously collected by Designers/Brand houses, stored in its central data-system such as electric point of sales systems (EPoS) or others generic database or spreadsheet softwares, and regularly updated, analysed, and utilised. Data can then be incorporated at the key 'integrated decision points' as illustrated in the 4 points of the Collective Activity Model (Figure 4). Furthermore, this information can be shared with long-term supply chain partners (e.g. textiles and garment manufacturers) as basis for new product development and/or quality performance improvement.

This intervention aids Designers/Brand houses' capability to anticipate the real and customised demand of the Customers, fine-tune their decision making for merchandise plan and make it more closely aligned with the Customers' needs and wants. It also in turn helps to reduce the errors associated with predicting some 12 months in advance and 'ordering in the dark', which often results in slow sell-through, costly mark-downs en mass, or large stock keeping. Furthermore, it creates a continuous 'feedback chain' for the collective processes between Producers and Customers through the FSC process, which is much more effective and timely than a long 'feedback loop'. It thus presents a direct contrast to the long feedback loop of the existing Fashion System practice, where information on customers is typically not gathered until the end of each season's sell-through cycle, and not incorporated into Producers designs until a minimum 3-4 months later, if at all. The new model therefore enhances the optimal balance between Producers' supply push and Customers' demand pull, by activating a MC implementation strategy at the heart of the FSC upstream processes.

On Production: Manufacturer Processes

Given that a) The prospect of mass customisation is indeed highly sought-after by the fashion Customers; and b) the validity and precision of design information passed through the FSC pipeline is of paramount importance in order to achieve the garment, as intended by the Producer and desired by the Customer; the field research investigated what and how textiles and garment manufacturers currently supply, and how they may re-organise existing processes in order to capitalise on this demand. The high occurrence and high cost of process reiteration, information duplication, and ad-hoc communication was identified throughout the FSC upstream processes, and is therefore a key area in the existing FSC practice that beckons streamlining.

New Model Strategy 2: Producers' Information Flow Re-Channeled

Fabric and garment specifications are identified as the critical design information generated by Designers and processed by Manufacturers all the way through the FSC pipeline to affect and/or determine the final garment sale to Customers. A set of unified fabric and garment specifications and their accompanying bill of material are thus proposed for designated functions between the FSC Producers, at particular communication interfaces in the total design process in the FSC upstream as illustrated in Figure 2. This ensures the correct content share at the right calendar time and between the right parties, and reduces the large margins for human error whiles reducing process delays thus streamlines process efficiency.

On Design: Designer Responds

On surveying international brand houses/designers companies, field research found that sample Designer/Brand-houses produce an average of some 800 garment styles per company per year, irrespective of the finer differences in each company's sub-season merchandising. The figures total the number of garment styles that actually sailed through the series of Producers' screening and selection processes in the lengthy FSC pipeline and end-up on the retail shelves. In other words, plenty more styles are designed, partially or fully developed, but dropped somewhere along the FSC pipeline. According to Eckert's research on ratio of design ideas to samples, only an approximate 30% of garment styles make it through as final merchandise (Eckert, 1999).

New Model Strategy 3: Designers' Information Creation Managed

As fabric, colour, and silhouettes are identified as the three key components of the garment object that repeatedly change in response to the parameters of a cyclical fashion system, it follows that using combinations and permutations of these three variables would offer an efficient strategy for the creation of new garment styles. A clear organisational and management structure of information classification and referencing can be devised to generate assembling options for the fashion garment, both in terms of design and production. This provides the creative and economic advantage for mass customised fashion garments and dramatically reduces complexity, time and resources wasted through repetitive process reiteration. Furthermore, it provides Producers with the ready capabilities to combine different design components, whist presenting Customers with a far more enticing product offering in the form of relevant options incorporating their individual requirements. The result is a far more mutually beneficial Prosumer relationship increasing both Customers' satisfaction and Producers' brand loyalty.

FURTHER RESEARCH

The proposal of this new model as an MC implementation strategy suitable for SMEs in the FSC, opens up new opportunities and possibilities for both Producers and Consumers with enhanced prospect for the synergy of 'Prosumers' co-design. It is noted that Burberry, amongst a handful high-end brand houses, has since this year taken the lead to take direct orders from Customers at the fashion week catwalk trade shows for goods delivery in 3-4 months time. DEPLOY demi-couture has also been pioneering in holding loyal Customers previews during collection design development processes, in order to collate Customers feedback for new season design finalisation and merchandising decisions. These new scenarios demonstrate that for worthy, desired fashion goods, Customers are indeed interested to engage in certain FSC processes and willing to wait for the necessary delivery lead time.

However, further study and testing remains to be done in this new arena of a mass customised design process, beyond the more widely explored aspects of sizing, made-to-measure, and enabling technology. Areas of key relevance are highlighted: 1) The in-depth investigation on the total design processes of FSC companies that have already adapted various types of MC strategies, 2) the precise content, mechanism, and system for translating Customer demands into sophisticated garment designs, and 3) the prospect of modularisation as a design management strategy for a mass customised FSC. Continuous research is also carried out on the optimisation matrix for collection design and development using mathematic permutations and combinations of the garment components. And whilst a challenging global economic conditions and market turbulence is faced by all, this climate will only serve to intensify the potential and necessity for variety and customisation (Pine, 1993; Harts, 1994), in order for Producers to fulfill Customers' demand more precisely.

REFERENCES

Alford, D., Sackett, P., & Nedler, G. (2000). Mass customization – an automotive perspective. *International Journal of Production Economics*, *65*, 99–110. doi:10.1016/S0925-5273(99)00093-6

Anderson-Connell, (2002). A consumer-driven model for mass customization in the apparel market. *Journal of Fashion Marketing and Management*, *6*(3). doi:10.1108/13612020210441346

Bardakci, A., & Whitelock, J. (2003). Mass-customisation in marketing: The consumer perspective. *Journal of Consumer Marketing*, *20*(5). doi:10.1108/07363760310489689

Blecker, T., & Abdelkafit, N. (2006). Modularity and delayed product differentiation in assemble-to-order systems. In T. Blecker & G. Freidrich (Eds.), *Mass customisation – Challenges and solutions*. New York, NY: Springer Science+Business Media, Inc.

Brannon, E. L. (2000). *Fashion forecasting*. New York, NY: Fairchild Publications.

Carr, H., & Pomeroy, J. (1992). *Fashion design and product development*. Oxford, UK: Blackwell Science.

Davis, S. (1996). *Future perfect*. Boston, MA: Addison-Wesley.

Easey, M. (Ed.). (1995). *Fashion marketing*. London, UK: Blackwell Science Ltd.

Eckert, C. (1999). Managing effective communication in knitwear design. *The Design Journal, 2*(3).

Feitzenger, E., & Lee, H. L. (1997). Mass customisation at Hewlett Packard: The power of postponement. *Harvard Business Review*, *75*(January-February).

Fiore, A. M. (2000). Relationships between optimum stimulation level and willingness to use mass customization options. *Journal of Fashion Marketing and Management*, *5*(2).

Forza, C., & Vinelli, A. (2000). Time compression in production and distribution within the textile-apparel chain. *Integrated Manufacturing Systems*, *11*(2). doi:10.1108/09576060010314134

Gaskill, L. R. (1992). Toward a model of retail product development – A case study analysis. *Clothing & Textiles Research Journal*, *10*(4), 17–24. doi:10.1177/0887302X9201000403

Gilmore II, J., & Pine II, B. J. (1997). The four faces of mass customization. *Clothing and Textile Research Journal, 17*.

Hart, C. W. L. (1995). Mass-customisation: Conceptual underpinnings, opportunities and limits. *International Journal of Service Industry Management*, *6*(2). doi:10.1108/09564239510084932

Jackson, T., & Shaw, D. (2001). *Fashion buying and merchandising management*. New York, NY: Palgrave.

Jones, R. M. (2002). *The apparel industry*. Oxford, UK: Blackwell Science.

Kay, M. J. (1993). *Making mass customisation happen: Lessons for implementation. Planning Review*. July-August.

Kotha, S. (1995). Mass customisation: Implementing the emerging paradigm for competitive advantage. *European Management Journal*, *14*(5).

Kotler, P. (1989). *From mass marketing to mass customisation. Planning Review*. September-October.

Lampel, J., & Mintzberg, H. (1996). Customising customisation. *Sloan Management Review*, *38*(1), 21–30.

Le Pechoux, et al. (2000). Developing a pattern language for fashion innovation management. *Journal of Fashion Marketing and Management*, *5*(4).

Lee, S.-E., & Chen, J. C. (1999). Mass-customization methodology for an apparel industry with a future. *Journal of Information Technology*, *16*(1).

Mc Kelvey, K., & Munslow, J. (2003). *Fashion design – Process, innovation & practice*. Oxford, UK: Blackwell Science.

Muffatto, M., & Roveda, M. (2000). Developing product platform: Analysis of the development process. *Technovation*, *20*, 617–630. doi:10.1016/S0166-4972(99)00178-9

Okonkwo, U. (2007). *Luxury fashion branding – Trends, tactics, techniques*. New York, NY: Palgrave Macmillan. doi:10.1057/9780230590885

Pan, B. (2006). A mass customised supply chain for the fashion system at the design-production interface. *Journal of Fashion Marketing and Management*, *10*(3). doi:10.1108/13612020610679312

Peppers, D., & Rogers, M. (1997). *The one to one future – Building relationships one customer at a time*. New York, NY: Currency Doubleday.

Piller, F. T. (2003). *Mass customisation – Ein wettbewerbsstrategisches Konzept im Informationszeitalter* (34th ed.). Wiesbaden, Germany: Gabler Verlag.

Pine, B. J. II. (1993). *Mass customisation: The new frontier in business competition*. Boston, MA: Harvard Business School Press.

Secor, L. (1992). *Computer usage in apparel design and its effect on styling and creativity*. Manchester: UMI.

Senanayake, M. M., & Little, T. J. (2010). Mass customization: Points and extent of apparel customization. *Journal of Fashion Marketing and Management*, *14*(2). doi:10.1108/13612021011046110

Sinha, P. (2001). The mechanics of fashion. In Hines, T., & Bruce, M. (Eds.), *Fashion marketing – Contemporary issues*. Oxford, UK: Butterworth-Heinemann.

Taylor, C. A. (2003). Implementing a mass customised clothing services. In Tseng, M. M., & Pillar, F. T. (Eds.), *The customer centric enterprises*. Berlin/Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-55460-5_26

Thackara, J. (1998). One-to-one online: Fashion, the Internet, and the role of design. In Malossi, G. (Ed.), *The style engine*. New York, NY: Monacelli Press, Inc.

Toffler, A. (1980). *The third wave*. New York, NY: William Morrow & Co. Inc.

Tseng, M. M., & Pillar, F. T. (2003). *The customer centric enterprises*. Berlin-Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-55460-5

Tsigkas, A. (2001). Distributed demand flow customisation. In Tseng, M. M., & Pillar, F. T. (Eds.), *The customer centric enterprises*. Berlin-Heidelberg, Germany: Springer-Verlag.

Westbrooks, R., & Williams, P. (1993). Mass customisation: Japan's new frontier. *European Journal of Management*, *11*(1).

Wickett, J. L., et al. (1999). Apparel retail product development: Model testing and expansion. *Clothing and Textile Research Journal, 1*.

Wilson, J. (2001). Predicting the future: An overview of the colour forecasting industry. *The Design Journal*, *4*(1), 15–30. doi:10.2752/146069201789378450

Zipkin, P. (2000). The limits of mass customization. *Sloan Management Review*, *42*(3), b1–b87.

This work was previously published in Fashion Supply Chain Management: Industry and Business Analysis, edited by Tsan-Ming Choi, pp. 251-269, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 16
Integration of Fuzzy Logic Techniques into DSS for Profitability Quantification in a Manufacturing Environment

Irraivan Elamvazuthi
Universiti Teknologi PETRONAS, Malaysia

Pandian Vasant
Universiti Teknologi PETRONAS, Malaysia

Timothy Ganesan
Universiti Teknologi PETRONAS, Malaysia

ABSTRACT

Production control, planning, and scheduling are forms of decision making, which play a crucial role in manufacturing industries. In the current competitive environment, effective decision-making has become a necessity for survival in the marketplace. This chapter provides insight into the issues relating to integration of fuzzy logic techniques into decision support systems for profitability quantification in a manufacturing environment. The chapter is divided into five sections with a general introduction of the topic, followed by a thorough literature review on the existing techniques. Thereafter, fuzzy logic algorithms using logistic membership functions and resource variables for decision making aiming at quality improvement are discussed. A case study involving a textile firm is then described with the computational results and findings, and finally, future research directions are presented.

INTRODUCTION

Decision making has become very important in several areas of human endeavor and in real-world situations, it involves information which is fuzzy in nature (Metaxiotis *et al*, 2003 and Monfared & Yan, 2004).However, most of our traditional tools for formal modeling, reasoning and computing are crisp, deterministic and precise in character. By crisp we mean dichotomous, which is, yes-or-no type rather than more-or-less type. In conventional dual logic, for instance, a statement can be true or false and nothing in between. In set theory,

DOI: 10.4018/978-1-4666-1945-6.ch016

an element can either belong to a set or not, and optimizing a solution is either feasible or not. Precision assumes the parameters of a model represent exactly either our perception of phenomenon modeled or the features of the real system that has been modeled (Sakawa & Yana, 1985).

A practical decision making problem is best tackled by mathematically modeling it. Often, a reasonable amount of mathematical models with incorporation of fuzziness are used in order to assist decision makers for making rational decisions. A discussion on fuzzy linear programming (FLP) with imprecision and uncertainty for decision making problems is provided. FLP handles fuzziness in the parameters where the model is parameterized to find a satisfactory solution in the form of function of membership values. This ensures easy computation (Delgado *et* al, 1989).

The continuous linear-programming involves decision variables that are fractional. However, when all decision variables are integers, it is known as integer-programming (Beasley, 2011 and Harvey, 2011),. A number of methods have been developed for integer programming such as enumeration, cutting-plane and group-theoretic techniques and they have been applied to various applications such as capital budgeting, warehouse location, scheduling, etc. (Bertsimas *et al*, 1999 and Ghanizadeh & Fakhri, 2009) However, often real-world problem, the realistic assumption is that does not always involve integers. That is the rational for this work where FLP method is used effectively for solving linear programs which involves non-integers. In this chapter, a methodology using logistic membership function to solve the FLP problem is presented.

BACKGROUND

The mathematical formalization of fuzziness was originally pioneered by (Zadeh, 1965). This paved the way for more research in the application of fuzzy in real-world problems (Orlovsky, 1980). Some earlier work on fuzzy decision making can be found in (Tamiz, 1996, Zimmermann, 1987, Kickert, 1978, Zimmermann, 1991). In the last decade, manufacturing companies decided to adopt intelligent solutions, since the traditional manufacturing decision-making mechanisms were found to be insufficiently flexible to respond to changing production styles and highly dynamic variations in product requirements (Kusiak, 1990 and Metaxiotis *et al*, 2002). (Custodio *et al,* 1994) discussed the issue of production planning and scheduling using a fuzzy decision system, whilst, several outlines concerning the development of a rule-base for the specification of manufacturing planning and control systems were made by (Howard *et al*, 2000). (Watada, 1997) has proposed one form of logistic membership function to overcome difficulties in using linear membership function in solving fuzzy decision making problem. Non linear logistic membership function was presented by (Vasant & Bhattacharya, 2007 and Bhattacharya & Vasant, 2007). Some representative publications can also be found in (Zimmermman, 1985, Yager *et al*, 1987, Dubois & Prade, 1980 Klir, & Yuan, 1995 and Ross, 1995)

The approach proposed here is based on interaction with the decision maker, the implementer and the analyst in order to find a compromised satisfactory solution for the fuzzy linear programming (FLP) problem. In a decision process using an FLP model, source resource variables may be fuzzy, instead of precisely given numbers in crisp linear programming (CLP) model. For example: machine hour, labor force, material needed and so on in a manufacturing center, are always imprecise, because of incomplete information and uncertainty in various potential suppliers and environments. Therefore, they should be considered as fuzzy resources, and FLP problem should be solved by using fuzzy set theory. In this chapter, a methodology to solve fuzzy linear programming problem with logistic membership is considered.

FUZZY SYSTEM

The heart of the methodology for the FLP problem as stated by (Rubin and Narasimhan, 1984) lies in the construction of membership function for the objection coefficients, technical coefficients, resource variables and decision variables. In this regards many researchers used the FLP approach in solving its applications problems in their work.

Various types of membership functions were used in fuzzy linear programming problem and its application such as a linear membership function, the linear fuzzy constraints, a tangent type of a membership function, an interval linear membership function, an exponential membership function, inverse tangent membership function, logistic type of membership function, concave piecewise linear membership function, piecewise linear membership function, flexible membership function, and dynamics membership function. As a tangent type of a membership function, an exponential membership function and hyperbolic membership function are nonlinear functions, a fuzzy mathematical programming defined with a nonlinear membership function results in nonlinear programming. Usually a linear membership function is employed in order to avoid nonlinearity. Nevertheless, there are some difficulties in selecting the solution of a problem written in a linear membership function. Therefore, in this chapter, a modified s-curve membership function is employed to overcome the deficiencies of linear membership function. Furthermore, S-curve membership function is flexible enough to describe the vagueness in the fuzzy parameters for the manufacturing problems (Vasant, 2005).

FUZZY MODEL

A linear programming problem is defined by (Elamvazuthi *et al*, 2010):

$$\begin{aligned}&\text{Maximize} \quad Cx\\&\text{Subject to} \quad Ax \le b, \quad x \ge 0.\end{aligned} \tag{1}$$

in which the components of a *1*x*n* vector *C*, an *m*x*n* matrix *A* and an *n*x1 vector *b* are all crisp parameters and *x* is an *n*-dimensional decision variable vector (Atanu *et al*,2008).

The system (1) may be redefined in a fuzzy environment with the following more elaborate structure:

$$\begin{aligned}&\text{Maximize} \quad \sum_{j=1}^{n} \tilde{c}_j x_j\\&\text{Subject to} \quad \sum_{j=1}^{n} \tilde{a}_{ij} x_j \underset{\sim}{\le} b_i, \quad i = 1, 2 \cdots m\end{aligned} \tag{2}$$

All fuzzy data $\tilde{c}_j \equiv \tilde{S}(c_j^a, c_j^b)$ and $\tilde{a}_{ij} \equiv \tilde{S}(a_{ij}^a, a_{ij}^b)$ are fuzzy variables with the following logistic membership functions (Kickert, 1978).,

$$\mu_{\tilde{c}_j} = \begin{cases} 1 & if\ c_j \le c_j^a \\ \dfrac{B}{1 + Ce^{\alpha\left(\frac{c_j - c_j^a}{c_j^b - c_j^a}\right)}} & if\ c_j^a \le c_j \le c_j^b \\ 0 & if\ c_j \ge c_j^b \end{cases} \tag{3}$$

$$\mu_{\tilde{a}_{ij}} = \begin{cases} 1 & if\ a_{ij} \le a_{ij}^a \\ \dfrac{B}{1 + Ce^{\alpha\left(\frac{a_{ij} - a_{ij}^a}{a_{ij}^b - a_{ij}^a}\right)}} & if\ a_{ij}^a \le a_{ij} \le a_{ij}^b \\ 0 & if\ a_{ij} \ge a_{ij}^b \end{cases} \tag{4}$$

where,

$$B = 1, C = 0.001, \alpha = 13.8$$

(Atanu et al, 2008).

The membership function of the resource variable b_i is given by

$$M_{bi} = \begin{cases} 1 & b_i \langle b_i^a \\ 0.999 & b_i = b_i^a \\ \dfrac{w}{1 + ue^{\left[\frac{b_i - b_i^a}{b_i^b - b_i^a}\right]}} & b_i^a \langle b_i \langle b_i^b \\ 0.001 & b_i = b_i^b \\ 0 & b_i \rangle b_i^b \end{cases} \quad (5)$$

where M_{bi} is the membership function of b_i and b_i^a, b_i^b the lower and the upper boundary of the fuzzy coefficient $\tilde{b}_i$ respectively.

$$\begin{aligned} M_{bi} &= \frac{w}{1 + ue^{\alpha\left[\frac{b_i - b_i^a}{b_i^b - b_i^a}\right]}} \\ e^{\alpha\left[\frac{b_i - b_i^a}{b_i^b - b_i^a}\right]} &= \frac{1}{u}\left[\frac{w}{M_{b_i}} - 1\right] \\ \alpha\left[\frac{b_i - b_i^a}{b_i^b - b_i^a}\right] &= Ln\frac{1}{u}\left[\frac{w}{M_{b_i}} - 1\right] \\ b_i &= b_i^a + \left[\frac{b_i^b - b_i^a}{\alpha}\right] Ln\frac{1}{u}\left[\frac{w}{M_{b_i}} - 1\right] \end{aligned} \quad (6)$$

Since b_i is a fuzzy coefficient for the resource variables, it is denoted as

$$\tilde{b}_i = b_i^a + \left[\frac{b_i^b - b_i^a}{\alpha}\right] Ln\frac{1}{u}\left[\frac{w}{M_{b_i}} - 1\right] \quad (7)$$

Figure 1 shows the membership function for resource variables.

The 'decision' in a fuzzy environment can be intersection of fuzzy objective functions. The relationship between constraints and objective function in fuzzy environment is fully symmetrical (i.e., no difference).

$$\tilde{D} = \tilde{C} \cap \tilde{G}$$

$$M_{\tilde{D}} = M_{\tilde{C}} \cap M_{\tilde{G}} \to Min\{M_{\tilde{C}} \cap M_{\tilde{G}}\}$$

In general, if we have (n) of objective function $(\tilde{G}_1, \tilde{G}_2, \ldots\ldots, \tilde{G}_n)$ and (m) $(\tilde{C}_1, \tilde{C}_2, \ldots\ldots, \tilde{C}_n)$

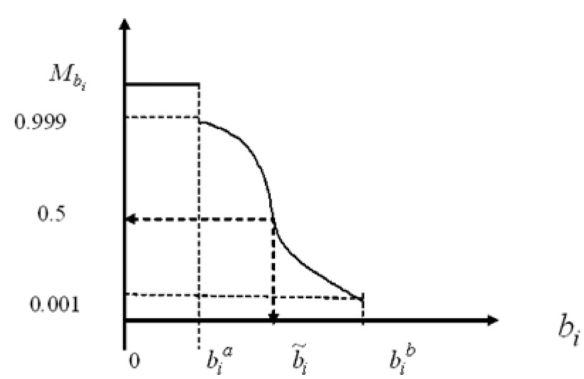
The decision

$$\begin{aligned} \tilde{D} &= \tilde{G}_1 \cap \tilde{G}_2, \ldots\ldots, \tilde{G}_n \cap \tilde{C}_1 \cap \tilde{C}_2, \ldots\ldots, \cap \tilde{C}_m \\ M_{\tilde{D}} &= \{Min\{M_{\tilde{G}_1}, M_{\tilde{G}_2}, \ldots\ldots, M_{\tilde{G}_n}\}, \\ & Min\{M_{\tilde{C}_1}, M_{\tilde{C}_2}, \ldots\ldots, M_{\tilde{G}_m}\}\} \\ M_{\tilde{D}} &= \\ & \{Min\{M_{\tilde{G}_1}, M_{\tilde{G}_2}, \ldots\ldots, M_{\tilde{G}_n}, M_{\tilde{C}_1}, M_{\tilde{C}_2}, \ldots\ldots, M_{\tilde{G}_m}\}\} \\ &= \{Min\{M_{\tilde{G}_j}, M_{\tilde{C}_i}\} \\ & i = 1, 2, \ldots\ldots, m \qquad j = 1, 2, \ldots\ldots, n \end{aligned} \quad (8)$$

$M_{\tilde{G}_j}$ = Degree of satisfaction for fuzzy objective function

$M_{\tilde{C}}$ = Degree of satisfaction for fuzzy constraints

Figure 1. Membership function for resource variables



$M_{\tilde{D}}$ = Degree of satisfaction for fuzzy decision

For example, the objective function "x should be substantially larger than 10" characterized by the membership function

$$M_{\tilde{G}}(x) = \begin{cases} 0 & x < 10 \\ (1+(x-10)^2)^{-1} & x \geq 10 \end{cases} \quad (9)$$

Constraint "x should be in the vicinity of 11, characterized by the membership function

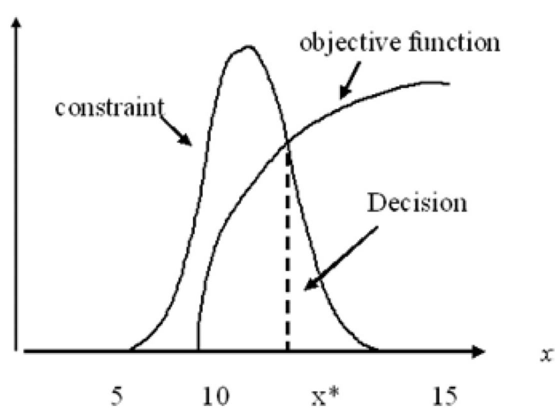
$$M_{\tilde{C}}(x) = (1+(x-1)^{-4})^{-1} \quad (10)$$

The membership function of the decision

$$M_{\tilde{D}}(x) = M_{\tilde{C}}(x) \cap M_{\tilde{G}}(x)$$
$$M_{\tilde{D}}(x) = \min\{(1+(x-10)^2)^{-1}, (1+(x-1)^{-4})^{-1}\}$$
$$M_{\tilde{D}}(x) = \begin{cases} 0 & x < 10 \\ (1+(x-10)^2)^{-1} & 10 \leq x \leq x^* \\ 1+(x-1)^{-4})^{-1} & x \leq x^* \end{cases} \quad (11)$$

Figure 2 shows the membership function of the decision.

Figure 2. Membership function of the decision



CASE STUDY

The data for this case study was derived from (Ertugrul & Tus, 2007). In this example the profit for a unit of sheet sales is around 1.05 Euro; a unit of pillow case sales is around 0.3 Euro and a unit of quilt sales is around 1.8 Euro. The projected sales are approximately 25.000 sheet units, 40.000 pillow case units and 10.000 units quilt units. The monthly working capacity and required process time for the production of sheets, pillow cases and quilts are given in Table 1 (Ertugrul & Tus, 2007).

Based on the Table 1, monthly production planning and profit for home-textile group is determined. x_1 presents the quantity of sheets that will be produced, x_2 presents the quantity of pillow cases and x_3 presents the quantity of quilts. The aspiration of the objective function is calculated by solving the following application data.

Application Data 1

If it is assumed that around 1.05 ≡ $\tilde{S}(1.02, 1.08)$, around 0.3 ≡ $\tilde{S}(0.2, 0.4)$, and around 1.8 ≡ $\tilde{S}(1.7, 2.0)$, then, the above problem can be modeled mathematically with fuzzy objective coefficients as follows Atanu *et al*, 2008):

Maximize

$$\tilde{S}(1.02, 1.08)x_1 + \tilde{S}(0.2, 0.4)x_2 + \tilde{S}(1.7, 2.0)x_3$$

subject to

$$\begin{aligned} &0.033x_1 + 0.01x_2 + 0.0033x_3 \leq 208; \\ &0.056x_1 + 0.25x_2 + 0.1x_3 \leq 4368; \\ &0.0067x_1 + 0.04x_2 + 0.17x_3 \leq 520; \\ &0.1x_1 + 0.1x_2 + 0.01x_3 \leq 780; \\ &x_1 \geq 25000; \\ &x_2 \geq 40000; \\ &x_3 \geq 10000; \end{aligned} \quad (12)$$

Table 1. Required process time for sheet, pillow case, and of a quilt (Ertugrul & Tus, 2007)

Departments	Required unit time(hour)			Working hours per month
	Sheet	Pillow case	Quilt	
Cutting	0.0033	0.001	0.0033	208
Sewing	0.056	0.025	0.1	4368
Pleating	0.0067	0.004	0.017	520
Packaging	0.01	0.01	0.01	780

Application Data 2

Maximize

$1.05x_1 + 0.3x_2 + 1.8x_3$ (profit)

Subject to

$$
\begin{aligned}
&0.0033x_1 + 0.001x_2 + 0.0033x_3 \leq (170, 208)\\
&0.0056x_1 + 0.025x_2 + 0.1x_3 \leq (3600, 4368)\\
&0.0067x_1 + 0.004x_2 + 0.0017x_3 \leq (520, 572)\\
&0.01x_1 + 0.01x_2 + 0.01x_3 \leq (780, 884)\\
&x_1 \geq 20,000\\
&x_2 \geq 40,000\\
&x_3 \geq 8,000\\
&x_1, x_2, x_3 \geq 0
\end{aligned} \tag{13}
$$

Simulation Algorithm

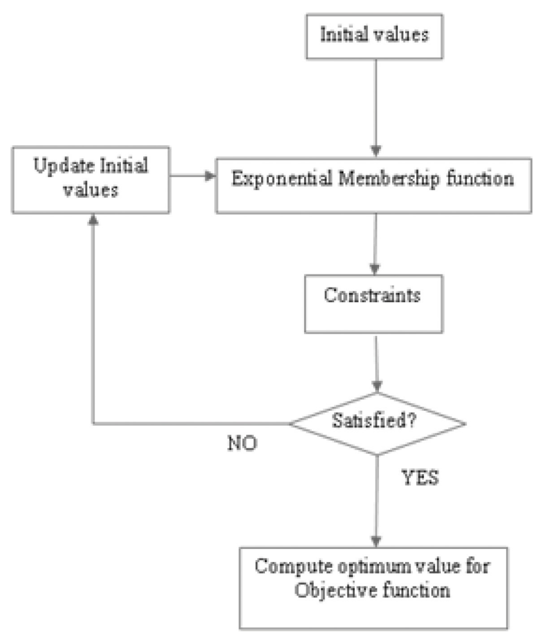
The fuzzy model in Equations (12) and (13) has been solved by using modified logistic membership function methodology and recursive iteration techniques. This was achieved through simulation based on the algorithm shown in Figure 3 (Elamvazuthi *et al*, 2010).

The algorithm for textile firm production planning is as shown below:

Step 1: set the initial values
Step 2: fuzzify using the exponential membership function
Step 3: incorporate the constraints
Step 4: if constraints satisfied, then, go to step 5
Step 5: compute the optimum value for the objective function output results
Step 6: else, re-establish initial values by using a recursive function with M = 748 (iteration, (Elamvazuthi *et al*, 2009)
Step 7: end

The numerical example is solved by using a recursive method for various iterations. This was carried out using the C++ programming language on a personal computer with a dual core processor running at 2 GHz (Intel Corporations, 2011 and Stroustrup B., 2011).

Figure 3. Algorithm



RESULTS AND DISCUSSION

The following section provides the computational and simulation results of the fuzzy model for application data 1 and 2 (Elamvazuthi *et al*, 2009 and Elamvazuthi *et al*, 2010).

Results of Application Data 1

Figure 4 shows the 3D outcome of the iteration with *M*=784 for various alpha values with respect to the objective function *G*. The values of α_1 and α_2 vary from 0 t o1. The optimum values for the objective function as per Figure 1 are 86,807.7 (maximum) and 86,755.4 (minimum).

Figure 5 shows the 3D outcome for *M* = 749 iterations and various alpha values with respect to *G*. The optimum values for the objective function as per this figure are 86,691.8 (maximum) and 86,639.5 (minimum).

Figure 6 shows the 3D outcome for *M* = 750 iterations and various alpha values with respect

Figure 4. 3D plot for iterations M=748

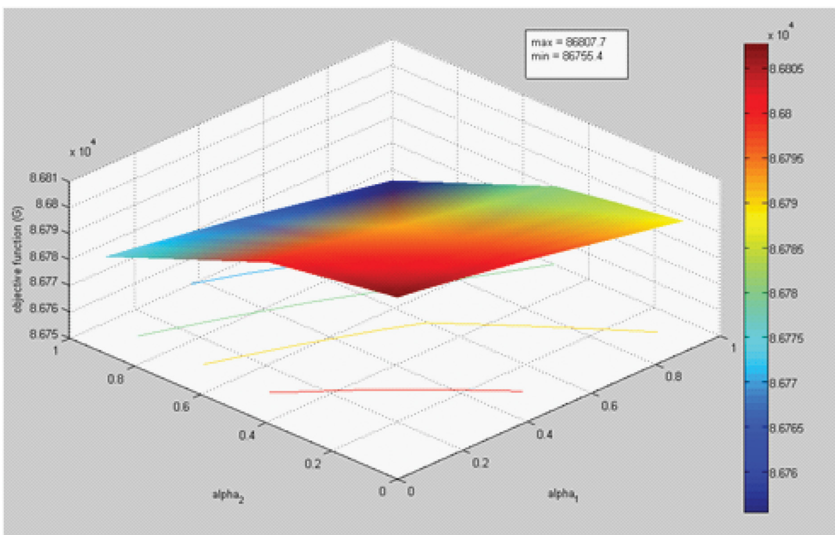
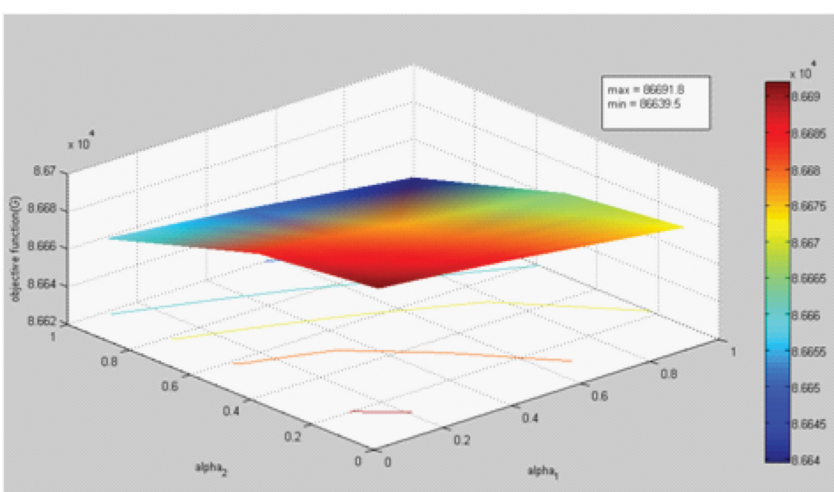


Figure 5. 3D plot for iterations M = 749



to *G*. The optimum values for the objective function as per this figure are 86,576.2 (maximum) and 86,524.0 (minimum).

Figure 7 shows the 3D outcome for $M = 751$ iterations and various alpha values with respect to *G*. The optimum value for the objective function as per this figure are 86,440.7 (maximum) and 86,408.0 (minimum).

Figure 8 shows the linear approximation for *G* with respect to iterations 748 to 751.

It can be seen that as the iterations are increased, the values of the objective function decrease. The percentage error is minimum at iteration, $M = 748$; however, after that it increases until it peaks at $M = 750$; thereafter, the percentage error decreases again to a level lower than that at $M = 748$. This shows that the maximum number of iterations that can be used for similar cases in the future can be limited to $M = 750$.

Figure 6. 3D plot for iterations M=750

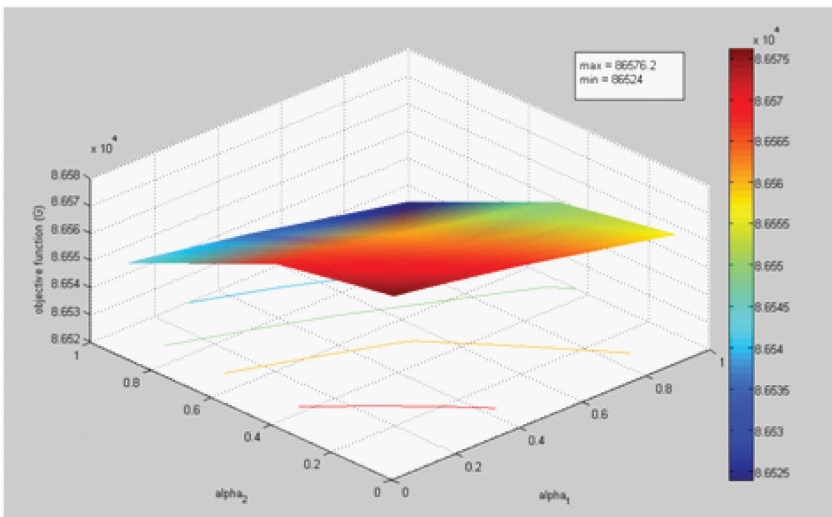


Figure 7. 3D plot for iterations M=751

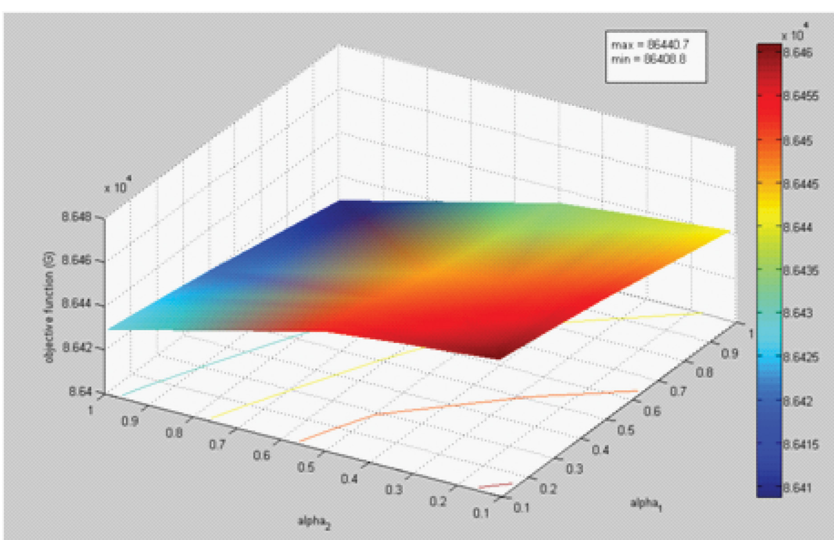


Figure 8. Objective function (G) versus iterations

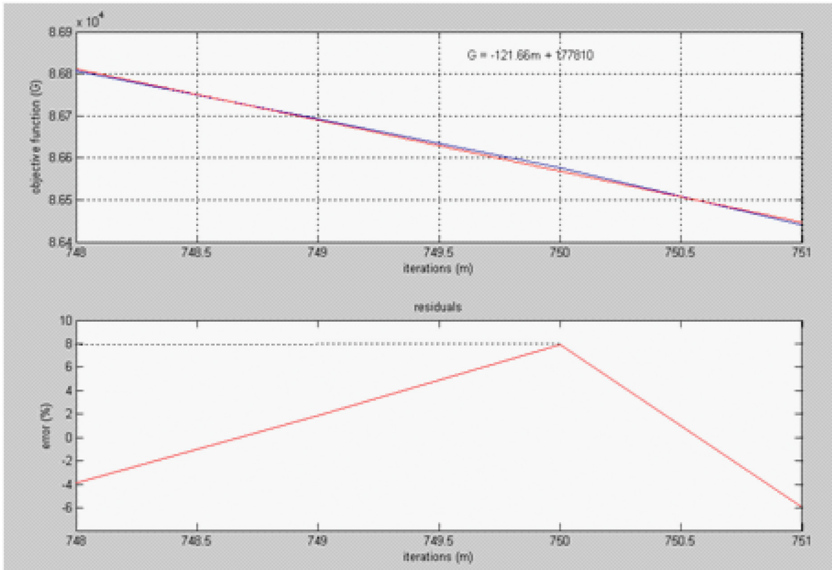


Figure 9, Figure 10, and Figure 11 show the linear approximation for the decision variables x_1, x_2 and x_3 with respect to the number of iterations.

It can be observed that x_1, x_2 and x_3 decrease as the iterations are increased from $M = 748$ to $M = 751$.

Table 2 presents results that involve α_1, α_2, and α_3 with $M = 748$ for G, x_1, x_2 and x_3.

Table 3 summarizes the result for $M = 748$ to 751 for x_1, x_2 and x_3 with maximum and minimum values of G. The overall maximum value for G is 86807.7 at $M = 748$ and the overall minimum value is 86408.0 at $M = 751$.

Table 4 compares the best objective function and decision variables x_1, x_2 and x_3 of the proposed method with previous work by other researchers.

Figure 9. Decision variable, X1 vs. M iterations

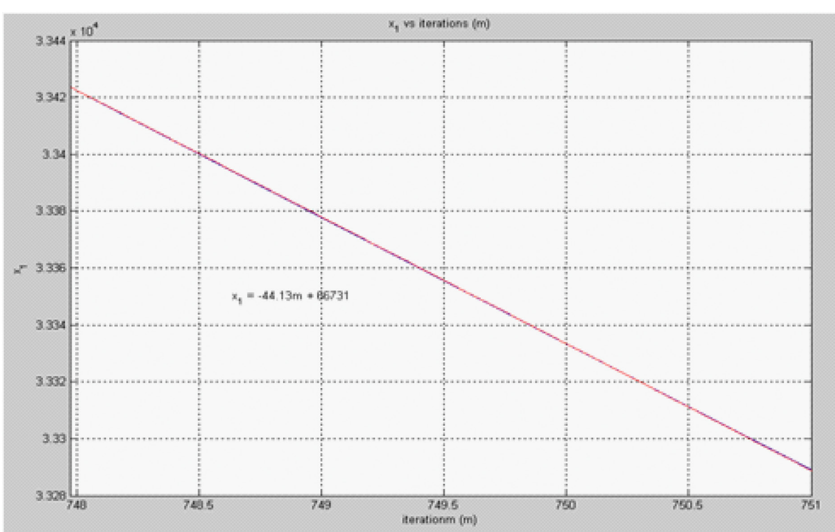


Figure 10. Decision variable, X2 vs. M iterations

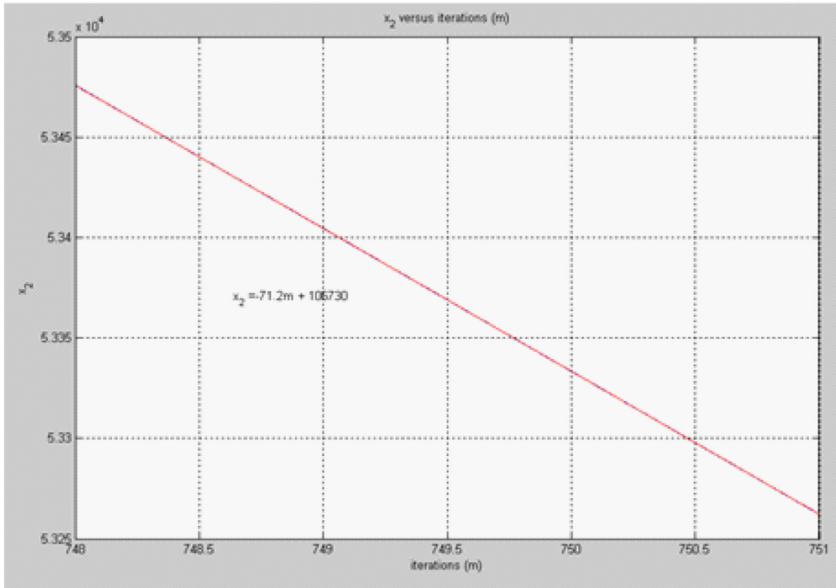
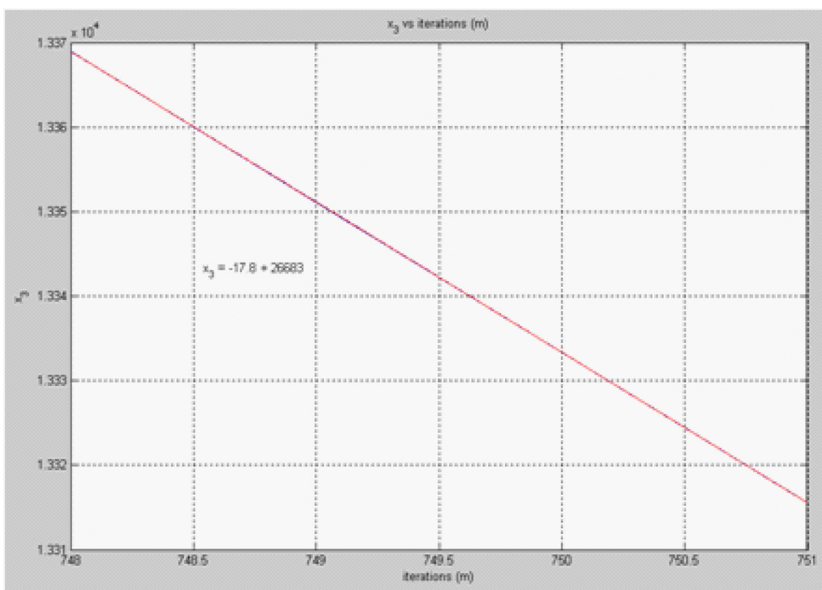


Figure 11. Decision variable, X2 vs. M iterations



From Table 4, the optimum value for the objective function using the proposed method outweighs the results obtained in (Klir & Yuan and 1995 and Ross, 1995). *Fuzzy Logic with Engineering Applications*, New York: McGraw- Hill.

It can be deduced that the recursive iteration method proposed here is an efficient and effective way to solve our example fuzzy problem of production planning in the textile industry.

Table 2. Alpha, objective function, and decision variables for M=748

α_1	α_2	α_3	G	x_1	x_2	x_3
1	1	*all	86755.4	33422.5	53475.9	13369
1	0.5	all	86780.3	33422.5	53475.9	13369
0.5	1	all	86767.5	33422.5	53475.9	13369
0.5	0.5	all	86792.4	33422.5	53475.9	13369
0.3333	1	all	86770.9	33422.5	53475.9	13369
0.3333	0.5	all	86795.9	33422.5	53475.9	13369
0.25	1	all	86772.5	33422.5	53475.9	13369
0.25	0.5	all	86797.5	33422.5	53475.9	13369
0.2	1	all	86773.5	33422.5	53475.9	13369
0.2	0.5	all	86798.4	33422.5	53475.9	13369
0.1667	1	all	86774.1	33422.5	53475.9	13369
0.1667	0.5	all	86799.0	33422.5	53475.9	13369
0.1429	1	all	86774.5	33422.5	53475.9	13369
0.1429	0.5	all	86799.4	33422.5	53475.9	13369
0.125	1	all	86774.8	33422.5	53475.9	13369
0.125	0.5	all	86799.8	33422.5	53475.9	13369
0.1111	1	all	86775.1	33422.5	53475.9	13369
0.1111	0.5	all	86800.0	33422.5	53475.9	13369

Note: *all∈ (0, 1)

Results of Application Data 2

Table 5 shows the results obtained for the objective function versus parameter d (Vagueness factor); Table 6 shows results obtained for the objective function versus θ (level of satisfaction), and Table 7 shows results obtained for the parameter d versus α(degree of possibility for three decision variables). The major contribution of this chapter is on the comparative results obtained on the objective function for various parameters d in (Ertugrul & Tus, 2007).

Figure 12 shows the relationship of α versus *d* parameter. One can precisely observe that the α value decreases as the value of d parameter increases.

Figure 13 depicts the 3D outcome the objective function, G versus θ versus *d*.

It can be seen that the objective function increases as the d parameter and θ are increased.

Figure 14 depicts the outcome the objective function, G_{max} versus θ.

It can be seen that the objective function increases as θ is increased.

Figures 15 to 23 show the outcome the objective function, versus d for various θ, ranging from 0.1 to 0.9. It can be seen that the objective function increases as θ is increased until it reaches steady state for the objective function.

For θ = 0.1, the G_{max} = 83715.2 and G_{min} = 83678.6. M is fixed at 748 iteration.
For θ =0.2, the G_{max} = 85530.9 and G_{min} = 85493.6. M is fixed at 748 iteration.
For θ =0.3, the G_{max} = 87325.3 and G_{min} = 87287.1. M is fixed at 748 iteration.
For θ =0.4, the G_{max} = 89196.5 and G_{min} = 89157.6. M is fixed at 748 iteration.
For θ =0.5, the G_{max} = 91039 and G_{min} = 90999.2. M is fixed at 748 iteration.

Table 3. Summary of iterations, decision variables, and objective function

M	x_1	x_2	x_3	G (max)	G(min)
748	33422.5	53475.9	13369	86807.7	86755.4
749	33377.8	53404.5	13351.1	86691.8	86639.5
750	33333.3	53333.3	13333.3	86576.2	86524.0
751	33288.9	53262.3	13315.6	86440.7	86408.0

Table 4. Comparative analysis

Method	The Best Objective Function	Decision Variables		
		x_1	x_2	x_3
(Ertugrul & Tus, 2007)	64390.999	33825.16	40000.00	9374.760
Atanu *et al*, 2008	66454.369	27766.99	40000.00	10233.01
Proposed Method	86807.700	33422.50	53475.90	13369.00

Table 5. Objective function vs. parameter D

				objective function (G)					
parameter d									
10	83678.6	85493.6	87287.1	89157.6	90999.2	92803.4	94680.6	96510.8	98283.9
15	83697.3	85512.6	87306.6	89177.5	91019.5	92824.1	94701.7	96532.3	98305.9
20	83711.3	85526.9	87321.2	89192.4	91034.8	92839.7	94717.6	96548.5	98322.3
25	83714.3	85530.0	87324.3	89195.6	91038	92843	94721	96551.9	98325.8
30	83715.0	85530.7	87325.1	89196.3	91038.8	92843.8	94721.8	96552.7	98326.7
35	83715.1	85530.9	87325.2	89196.5	91038.9	92843.9	94721.9	96552.9	98326.8
40	83715.2	85530.9	87325.3	89196.5	91039	92844	94722	96552.9	98326.9
45	83715.2	85530.9	87325.3	89196.5	91039	92844	94722	96552.9	98326.9
50	83715.2	85530.9	87325.3	89196.5	91039	92844	94722	96552.9	98326.9
θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
G_{max}	83715.2	85530.9	87325.3	89196.5	91039	92844	94722	96552.9	98326.9
G_{min} (Ertugrul & Tus, 2007)	83678.6 60538.0	85493.6 61501.0	87287.1 -	89157.6 -	90999.2 -	92803.4 -	94680.6 -	96510.8 -	98283.9 67281.0

Table 6. Objective function vs. θ

θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
G_{max}	83715.2	85530.9	87325.3	89196.5	91039	92844	94722	96552.9	98326.9
G_{min} (Ertugrul & Tus, 2007)	83678.6 60538.0	85493.6 61501.0	87287.1 -	89157.6 -	90999.2 -	92803.4 -	94680.6 -	96510.8 -	98283.9 67281.0

Table 7. Parameter d vs. α

d	α_1	α_2	α_3
10	0.870767	0.870767	0.972733
15	0.35612	0.35612	0.870767
20	0.043428	0.043428	0.559982
25	0.003713	0.003713	0.193789
30	0.000305	0.000305	0.043428
35	2.51E-05	2.51E-05	0.008502
40	2.06E-06	2.06E-06	0.00162
45	1.69E-07	1.69E-07	0.00031
50	1.39E-08	1.39E-08	5.78E-05

For θ =0.6, the G_{max} = 92844 and G_{min} = 92803.4. *M* is fixed at 748 iteration.
For θ =0.7, the G_{max} = 94722 and G_{min} = 94680.6. *M* is fixed at 748 iteration.
For θ =0.8, the G_{max} = 96552.9 and G_{min} = 96510.8. *M* is fixed at 748 iteration.
For θ =0.9, the G_{max} = 98326.9 and G_{min} = 98283.9. *M* is fixed at 748 iteration.

The best results obtained for various parameter *d* (Vagueness factor) respect to level of satisfaction θ has been compared with (Ertugrul & Tus, 2007). The findings reveal that the proposed method outweighs compared with (Ertugrul &

Figure 12. α vs. d

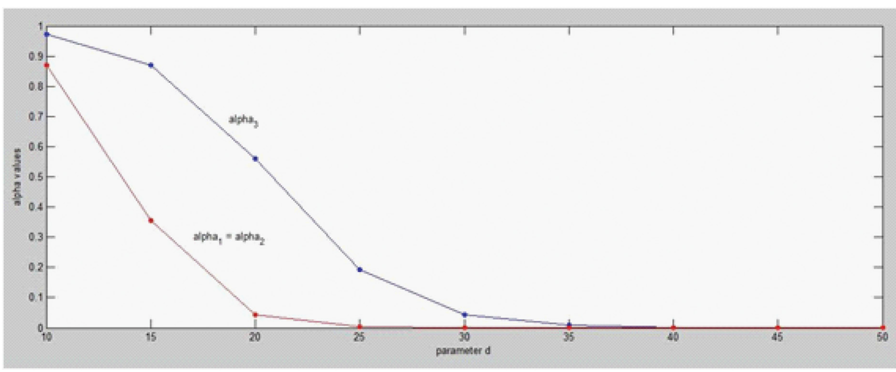


Figure 13. Objective function, G vs. θ versus d

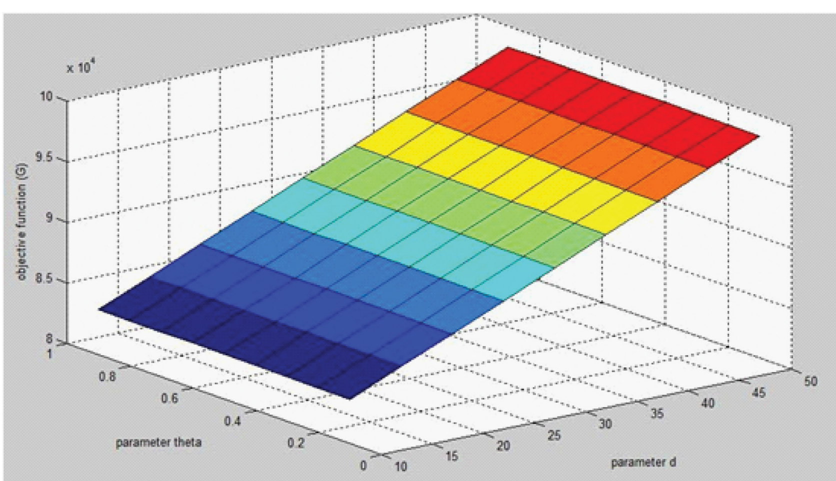


Figure 14. Objective function, G_{max} vs. θ

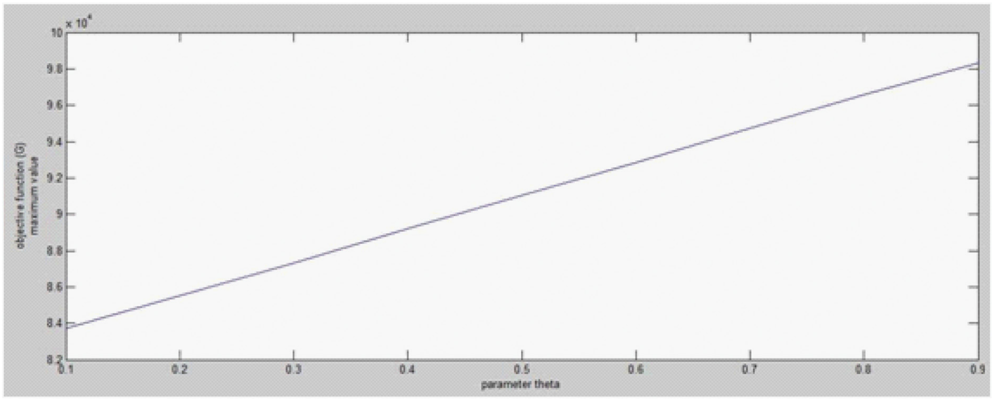


Figure 15. Objective function, G vs. d for θ =0.1

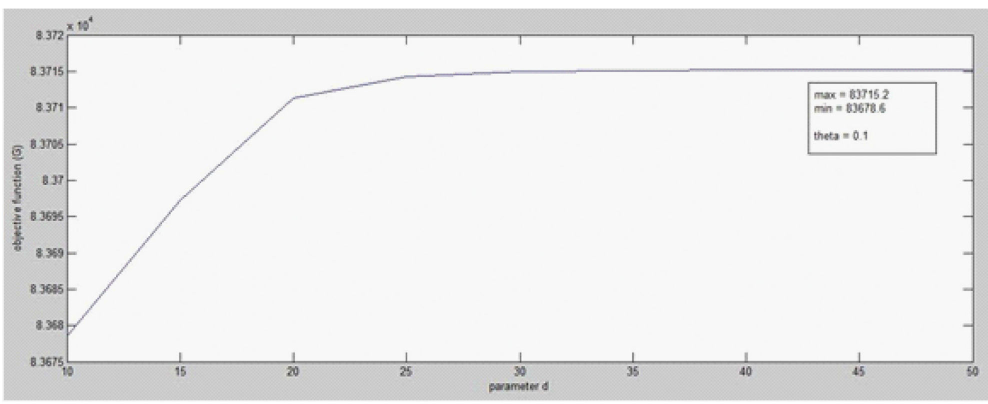


Figure 16. Objective function, G vs. d for θ = 0.2

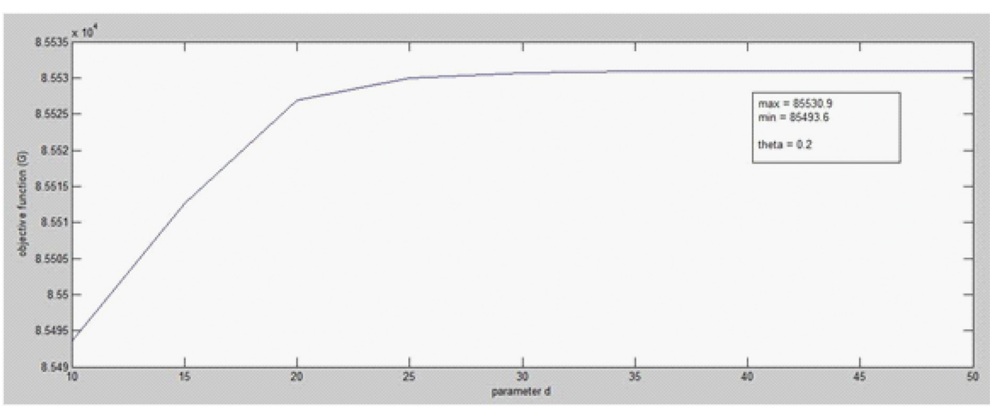


Figure 17. Objective function, G vs. d for θ = 0.3

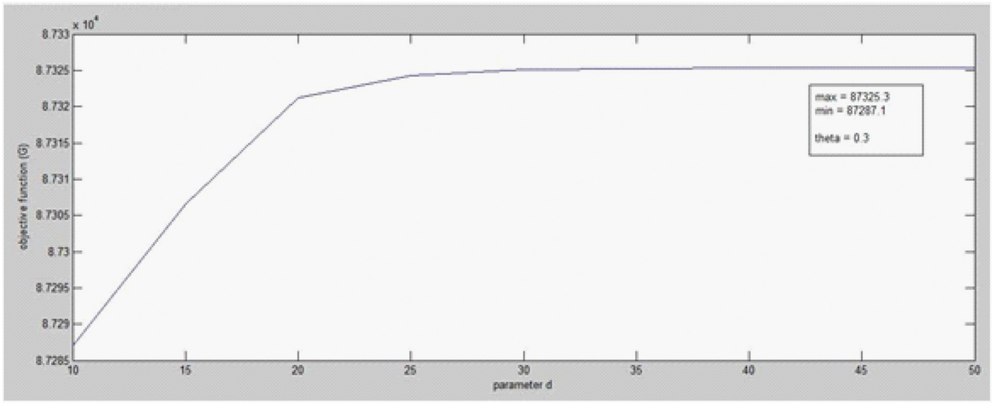


Figure 18. Objective function, G vs. d for θ=0.4

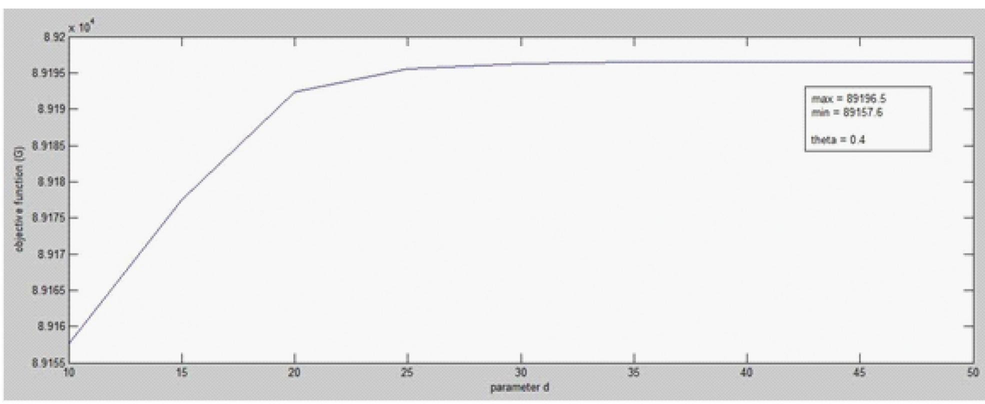


Figure 19. Objective function, G vs. d for θ =0.5

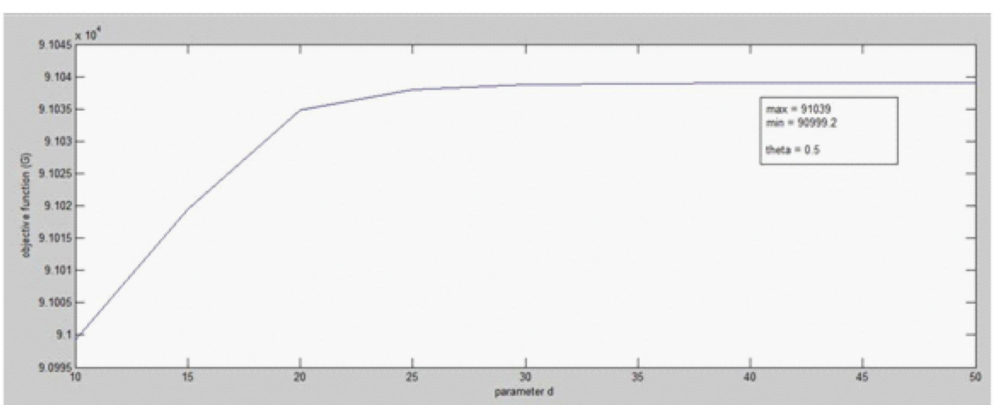


Figure 20. Objective function, G vs. d for θ = 0.6

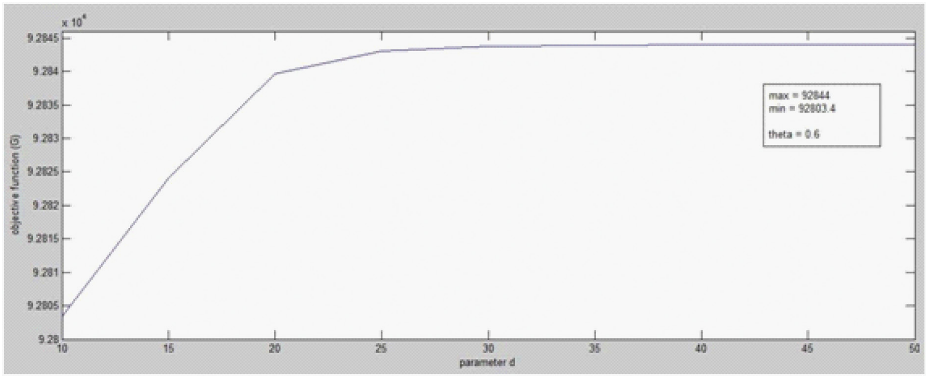


Figure 21. Objective function, G vs. d for θ =0.7

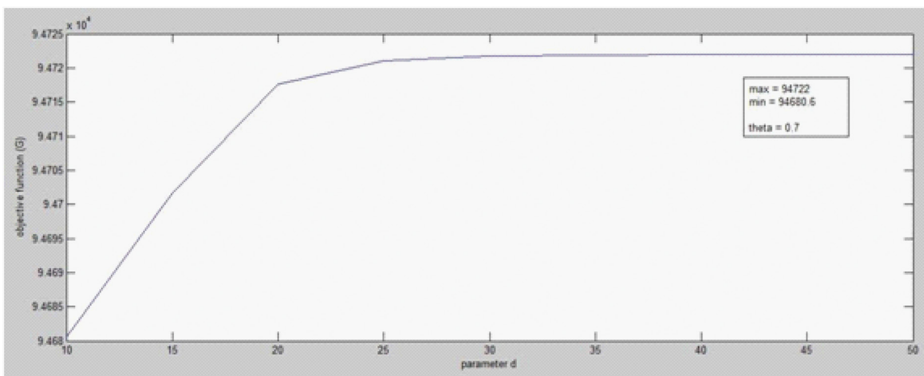


Figure 22. Objective function, G vs. d for θ =0.8

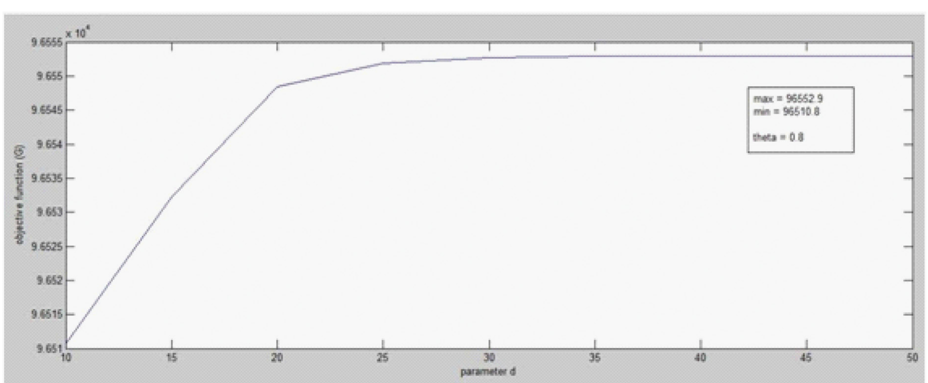
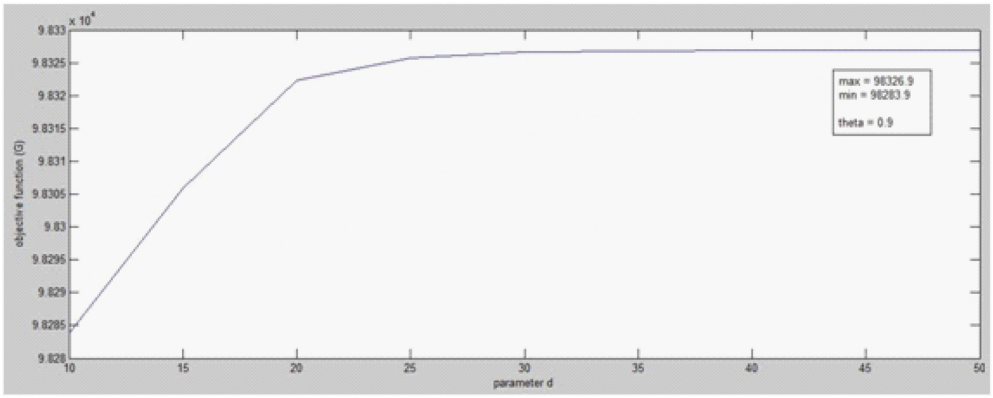


Figure 23. Objective function, G vs. d for θ =0.9



Tus, 2007). The authors strongly believe that the proposed iterative method on decision variables can be possibly applied to other large scale industrial production problems in real life applications. The major advantages of the proposed method is on the computational time in finding the objective function as well as the strength of the method on obtaining a very high feasible solution for the decision variables. The authors have put more emphasis on the well developed algorithms on the particular interest for the increasing the values for the decision variables.

FUTURE RESEARCH DIRECTIONS

In the future, hybrid optimization techniques can be applied to this industrial production planning problem in the textiles firm. Genetic algorithms, simulated annealing, tabu search and other heuristics and meta-heuristics techniques will be a very significant approaches in order to provide a satisfactory optimal solutions(Zhu & Zou, 2005, Ramirez *et al* 2009, Dawei & Shiyuan, 2007, Sharma & Lees, 2004, Sun *et al*, 2007 and Ismail, 2008).

CONCLUSION

In this chapter, the fuzzy model with fuzzy resource variables is successfully solved by fuzzy linear programming method with modified S-curve membership function. The recursive iterative technique has been intensively applied on the optimal solution for the objective profit function and feasible solutions for the decision variables. It can be seen that objective function using the proposed method produces a maximum value which better than previous research results. Furthermore the computational time (CPU) for the current method, in finding the optimal solution, is less than one second. Therefore the current method adopted in this chapter is very flexible, robust and importantly it is highly productive. We can observe that fuzzy linear programming provides more information to the decision maker and implementer to solve practical decision making problems in the application area of textiles industry. This modern artificial intelligence approach is a very useful tool in the decision making process when the environment is under turbulence and unpredictable uncertain.

REFERENCES

Atanu, S., Vasant, P., & Cvetko, J. A. (2008). Fuzzy optimization with robust logistic membership function: A case study in for home textile industry. *Proceedings of the 17th World Congress, The International Federation of Automatic Control*, Seoul, Korea, July6-11, (pp. 5262-5266).

Bertsimas, D., Darnell, C., & Soucy, R. (1999). Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company. *Interfaces*, *29*, 49–66. doi:10.1287/inte.29.1.49

Bhattacharya, A., & Vasant, P. (2007). Soft-sensing of level of satisfaction in TOC product-mix decision heuristic using robust fuzzy-LP. *European Journal of Operational Research*, *177*, 55–70. doi:10.1016/j.ejor.2005.11.017

Custodio, L. M. M., Sentieiro, J. J. S., & Bispo, C. F. G. (1994). Production planning and scheduling using a fuzzy decision system. *IEEE Transactions on Robotics and Automation*, *10*(2), 160–168. doi:10.1109/70.282541

Dawei, J., & Shiyuan, W. (2007). *Using the simulated annealing algorithm for multiagent decision making. RoboCup 2006, LNAI 4434* (pp. 110–121). Berlin, Germany: Springer-Verlag.

Delgado, M., Verdegay, J. L., & Vila, M. A. (1989). A general model for fuzzy linear programming. *Fuzzy Sets and Systems*, *29*, 21–29. doi:10.1016/0165-0114(89)90133-4

Dubois, D., & Prade, H. (1980). *Fuzzy sets and systems, theory and applications*. Academic Press Inc.

Elamvazuthi, I., Ganesan, T., Vasant, P., & Webb, J. F. (2009). Application of a fuzzy programming technique to production planning in the textile industry. *International Journal of Computer Science and Information Security*, *6*(3), 238–243.

Elamvazuthi, I., Vasant, P., & Ganesan, T. (2010). Fuzzy linear programming using modified logistic membership function. [Theory and Applications]. *International Review of Automatic Control*, *3*(4), 370–377.

Ertugrul, I., & Tus, A. (2007). Interactive fuzzy linear programming and an application sample at a textile firm. *Fuzzy Optimization and Decision Making*, *6*, 29–49. doi:10.1007/s10700-006-0023-y

Ghanizadeh, A. R., & Fakhri, M. (2009). A mixed-integer programming model to determine the optimum design of flexible pavement structure. *Australian Journal of Basic and Applied Sciences*, *3*(3), 1652–1660.

Howard, A., Kochhar, A., & Dilworth, J. (2000). Case studies based development of a rulebase for the specification of manufacturing planning and control systems. *International Journal of Production Research*, *38*(12), 591–606. doi:10.1080/002075400411376

Ismail, Z. (2008). A decision support system for improving forecast using genetic algorithm and Tabu search. *ARPN Journal of Engineering and Applied Sciences*, *3*(3), 13–16.

Kickert, W. J. (1978). *Fuzzy theories on decision-making: Frontiers in systems research*. Leiden, The Netherlands: Martinus Nijhoff.

Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. Upper Saddle River, NJ: Prentice Hall PTR.

Kusiak, A. (1990). *Intelligent manufacturing systems*. Englewood Cliffs, NJ: Prentice-Hall.

Metaxiotis, K., Psarras, J., & Samouilidis, E. (2003). Integrating fuzzy logic into decision support systems: current research and future prospects. *Information Management & Computer Security*, *11*(2), 53–59. doi:10.1108/09685220310468592

Metaxiotis, K., Samouilidis, E., & Psarras, J. (2002). Artificial intelligence in production scheduling: two decades of research and promises, invited chapter. *International Symposium on Scheduling*, Hamamatsu Japan, (pp. 5-12).

Monfared, M. A. S., & Yan, J. B. (2004). Design of an intelligent manufacturing scheduling and control system using fuzzy logic: Sensitivity analysis and parameter optimization. *Journal of Intelligent and Fuzzy Systems*, *15*(2).

Orlovsky, S. A. (1980). On formalization of a general fuzzy mathematical programming problem. *Fuzzy Sets and Systems*, *3*, 311–321. doi:10.1016/0165-0114(80)90026-3

Ramirez, A. J., Knoester, D. B., Cheng, B. H. C., & McKinley, P. K. (2009). *Applying genetic algorithms to decision making in autonomic computing systems.* ICAC'09, Barcelona, Spain.

Ross, T. J. (1995). *Fuzzy logic with engineering applications*. New York, NY: McGraw- Hill.

Rubin, P. A., & Narasimhan, R. (1984). Fuzzy goal programming with nested priorities. *Fuzzy Sets and Systems*, *14*, 115–129. doi:10.1016/0165-0114(84)90095-2

Sakawa, M., & Yana, H. (1985). Interactive decision making for multi-objective linear fractional programming problems with fuzzy parameters. *Cybernetics and Systems*, *16*, 377–397. doi:10.1080/01969728508927781

Sharma, S. K., & Lees, B. G. (2004). *A comparison of simulated annealing and GIS based MOLA for solving the problem of multi-objective land use assessment and allocation. MCDM, Whistler*. B. C. Canada.

Sun, X., Huang, M., & Wang, X. W. (2007). *Tabu search based distributed risk management for virtual enterprise.* Second IEEE Conference on Industrial Electronics and Applications.

Tamiz, M. (1996). *Multi-objective programming and goal programming: Theories and applications*. Berlin, Germany: Springer-Verlag.

Vasant, P. (2005). Solving linear programming with modified S-curve membership function. *International Journal of Uncertainty*. *Fuzziness and Knowledge-Based System*, *13*(1), 97–109. doi:10.1142/S0218488505003321

Vasant, P., & Bhattacharya, A. (2007). Sensing degree of fuzziness in MCDM model using modified flexible S-curve MF. *International Journal of Systems Science*, *38*(4), 279–291. doi:10.1080/00207720601117108

Watada, J. (1997). Fuzzy portfolio selection and its applications to decision making. *Tatra Mountains Mathematics Publication*, *13*, 219–248.

Yager, R. R., Ovchinikov, S., Tong, R. M., & Nguyen, H. T. (Eds.). (1987). *Fuzzy sets and applications-Selected chapters by L. A. Zadeh*. New York, NY: John Wiley.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353. doi:10.1016/S0019-9958(65)90241-X

Zhu, L. X., & Zou, L. (2005). Application of genetic algorithm in decision-making optimization of underground gas pipeline risk model. *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, China.

Zimmermann, H. J. (1987). *Fuzzy sets, decision making, and experts systems*. Boston, MA: Kluwer. doi:10.1007/978-94-009-3249-4

Zimmermann, H. J. (1991). *Fuzzy set theory-And its applications* (2nd rev. ed.). Boston, MA: Kluwer.

Zimmermman, H. J. (1985). Application of fuzzy set theory to mathematical programming. *Information Sciences*, *36*, 25–58. doi:10.1016/0020-0255(85)90025-8

ADDITIONAL READING

Araghi, A. R., & Abedi, M. (2009). An application of fuzzy tuning and TCPS for enhancement of dynamic performance of two-area reheat thermal system. *International Review of Automatic Control*, *2*(6), 725–730.

Bazaraa, M. S., Jarvis, J. J., & Sherali, H. D. (1990). *Linear programming and network flows*. John Wiley & Sons.

Caballero, R., Laguna, M., Martı, R., & Molina, J. (2010). Scatter Tabu search for multiobjective clustering problems. *The Journal of the Operational Research Society*, *62*(11), 1–13.

David, D., Olivier, R., & Arnauld, H. (2010). Comparison of different Tabu search operators on a highly constrained industrial problem. *International Journal of Advanced Operations Management*, *1*(4), 363–379.

Dedekova, Z., & Kozak, S. (2009). Modeling and control of non-linear systems using soft computing methods. *International Review of Automatic Control*, *2*, 366–371.

Herrera, F., & Verdegay, J. L. (1991). *Approaching fuzzy linear programming problems, interactive fuzzy optimization*. Springer-Verlag.

Lai, Y. J., & Hwang, C. L. (1992). *Fuzzy mathematical programming. Lecture Notes in Economics and Mathematical Systems* (*Vol. 164*). Springer-Verlag.

Roberto, S., Oscar, M. R., Oscar, C., & Patricia, M. (2010). Optimizing the MFs in type-2 fuzzy logic controllers, using the human evolutionary model. *International Review of Automatic Control*, *3*(1), 1–10.

Taher, S. A., & Amooshi, M. K. (2010). Optimal load frequency control using PSO algorithms in deregulated power systems. *International Review of Automatic Control*, *2*, 708–714.

Tanaka, H., & Asai, K. (1984). Fuzzy linear programming problems with fuzzy numbers. *Fuzzy Sets and Systems*, *13*, 1–10. doi:10.1016/0165-0114(84)90022-8

KEY TERMS AND DEFINITIONS

Decision Support System: A method to model data and make quality decisions based on it to support business or organizational decision-making activities

Fuzzy Linear Programming: An application of fuzzy set theory in linear decision making problems to determine an optimal solution by considering a number of constraints

Fuzzy Set Theory: Set membership that handles the concept of partial truth where the truth values fall between completely true and completely false

Logistic Membership Function: An appropriate function to represent the degree of truth

Manufacturing Environment: An industrial operational facility to fabricate products

Objective Function: A function that can be made as large or as small as possible

Optimal Value: Best state that a value could achieve

Quality Improvement: Systematic approach to improve product / service level compared with accepted standards

Resource Variable: A resource is used to represent dynamic data that holds as a reference to an external resource

This work was previously published in Handbook of Research on Industrial Informatics and Manufacturing Intelligence: Innovations and Solutions, edited by Mohammad Ayoub Khan and Abdul Quaiyum Ansari, pp. 171-192, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 17
Control Model for Intelligent and Demand-Driven Supply Chains[1]

Jan Ola Strandhagen
SINTEF Technology and Society, Norway

Heidi Carin Dreyer
Norwegian University of Science and Technology, Norway

Anita Romsdal
Norwegian University of Science and Technology, Norway

ABSTRACT

Orchestrating supply chains is challenging. This chapter describes how to control a supply chain to make it truly demand-driven – based on the assumption that all relevant information is made available to all partners in real time. The chapter explores the elements of a framework for intelligent and demand-driven supply chain control, with regards to the overall concept and associated principles, and demonstrates these in a case example. Challenges to the realization of the proposed control model include trust and power, supply chain dynamicity and uncertainty, and required investments in competence, standardization, and information and communication technology. Some of these can be met through initial small-scale implementations of the proposed model, to demonstrate effects, and by exploiting facilities for information sharing and collaboration, like supply chain dashboards and control studios. Future research within operations management, technology and information and communications technology (ICT) will support broader realization of the proposed control model.

INTRODUCTION

Integrating and coordinating supply chain and network operations are considered prerequisites for achieving high efficiency and competitiveness. Focusing on the performance and competitiveness of the supply chain, rather than the single company is a trend in several industries. Increased outsourcing and globalization greatly expands the complexity of supply chain operations and planning and control processes. A supply chain often has a decentralized geographical structure with complex logistics, where the various organizational units have a high degree of autonomy.

DOI: 10.4018/978-1-4666-1945-6.ch017

Thus, supply chain operations require a well-defined orchestration of a broad set of activities, resources and companies, with the common goal of fulfilling the demand of the end customer.

Designing supply chains is challenging due to their heterogeneous system characteristics, diversified product and material flow structure, trade-off situations, and conflicting interests and goals of the participants. Products vary in value, volumes and shelf life. Today, offering high customer service means either maintaining a high stock level or having frequent deliveries. Conflicts of interest often arise when a manufacturer aims to utilize economies of scale by producing large volumes and buffering products in stock, only to discover that there is end customer demand only for a minor part of the volume. Meanwhile, the wholesaler strives to keep stock levels low and buy only according to end customer demand.

A number of challenges related to supply chain and network operations can be identified. First, operations of supply chains require development of a unified set of control principles which simultaneously harmonize the interests of the involved companies and the entire network. These control principles should also adapt operations to end customer demand, and provide for the right balance of cost, time and service performance measures for each product and the supply chain as a whole. Second, a number of technical challenges are related to the information and communication technology (ICT) solutions that are required to provide supply chain members with access to real-time information on network operations. This includes data capturing technology, communication platforms, visualization technology, and decision support tools which require solutions to standardization issues and considerable investments. Third, the aspects of trust and companies' lack of willingness to share market and demand information is an obstacle to achieving demand-driven collaboration. Many companies perceive demand information to be a key element in protecting their own power position and staying competitive.

The main focus in the work described in this chapter is the interplay of the first two challenges; *how to control a supply chain to make it truly demand-driven,* based on the assumption that all relevant information is made available to all partners in real time. The objective is therefore to explore the elements of a framework for intelligent and demand-driven supply chain control with regards to the overall concept and associated principles, and to demonstrate these in the case of a supply chain for processed salads.

In this chapter, a concept and a number of associated principles are illustrated, in the form of a control model based on the use of real-time information. The chapter contributes to theory and knowledge through its description of an integrated framework for supply chain control and associated principles and the demonstration of these in a practical case. The aim of the framework and principles is to assist managers of supply chain organizations by highlighting and exploring possibilities and some key issues related to control of supply chain operations. A further contribution to practice is derived through the exemplification of how the framework and proposed control model can be applied in an actual supply chain.

The work is based on the *control model methodology* (Alfnes, 2005; Alfnes & Strandhagen, 2000). Viewing network and supply chain operations from a control model perspective is a systematic approach that enables integration and network partnerships for increased competitiveness. This methodology has been applied in numerous industrial cases over the past decades, with the experiences and results serving as a practical platform for a deductive research approach. This action-based, reengineering type of methodology, focusing on visual presentation and communication between involved supply chain members, also enables them to meet the challenges related to lack of trust and understanding between supply chain partners.

BACKGROUND

The supply chain approach is assumed to be a winning strategy for improving competitiveness and conducting business development. A vast amount of literature shows how a holistic and integrated supply chain strategy can lead to competitive advantages and other synergy effects, such as increased trust and less complex contract mechanisms between business partners, increased service levels, reduced costs and increased economies of scale (see e.g. Bowersox, Closs, & Stank, 1999; Christopher, 1998; Min & Mentzer, 2004). However, the success of a supply chain strategy depends on a number of factors, such as how well the supply chain as a whole is operated and controlled, to what extent operations are in line with demand and customer requirements (Simchi-Levi, Kaminsky, & Simchi-Levi, 2008; Slack, Chambers, & Johnston, 2001; Vollmann, Berry, Whybark, & Jacobs, 2005), how responsive the chain is to changes and uncertainty (Reichart & Holweg, 2006), and whether the companies involved are able to collaborate as if they where a unified entity (Tyndall, Gopal, Partsch, & Kamauff, 1998). In addition, ICT provides unprecedented potential as enabler of new ways of operating the supply chain and for companies to interact and share information (Clark & Lee, 2000; Ellinger, Taylor, & Daugherty, 1999).

Supply chain operations extend the control span and increase the complexity of the planning and control task. The increasing need for coordination, integration and collaboration among organizations has led to a wide range of business initiatives and projects that seek to investigate the potential of supply chain strategies, solutions and developments this requires, and how this will affect existing business models and relationships. Some relevant examples from past and ongoing research and development (R&D) projects are described below.

The pharmaceutical industry is characterized by a wide variety of products, both prescription and non-prescription drugs, with limited shelf life. These products are offered through a supply chain that experiences continuing pressure for reducing costs, lead times and stock levels, while simultaneously increasing turnover rates and product availability. A significant challenge for pharmaceutical supply chains is how to achieve improvements and shorten the time from production to market. The R&D project Automed (2004-2007, see www.sintef.no/automed) focused on the supply chain from a pharmaceutical production plant, through a pharmaceutical supplier and a wholesale unit, to a chain of Norwegian pharmacies. Analysis found that supply chain performance could be significantly improved, for instance, that cost levels could be reduced by 25-30%, lead time could be halved and stock levels reduced by 30%. In order to realize this potential, the involved companies had to change their entire supply chain approach towards tighter integration and application of demand-driven, pull-based control principles based on sharing of point-of-sales (POS) data. Part of the potential was realized through the implementation of an automated replenishment solution, where the responsibility for pharmacy replenishment was shifted backwards in the chain to the wholesaler.

A study of the grocery sector shows similar challenges. In a project called Smart Vareflyt, or Smart Flow of Goods (2007-2009, see www.sintef.no/smartvareflyt), nine supply chain members in the Norwegian grocery industry (fresh food producers, packaging producers, grocery wholesalers and retailers, and logistics service providers) cooperated, in order to investigate how real-time information would affect the flow of goods and information and what effects sharing of demand information in the supply chain could have on performance. Reducing supply chain lead time and increasing shelf life for perishable products in stores is a high priority issue in the grocery industry. A main objective of the project was to test and demonstrate the potential of radio frequency identification (RFID) of products and packaging and application of electronic product code information services (EPCIS) for information sharing. Analysis showed that the various supply chains could improve their performance if the traditional make-to-stock (MTS) approach were replaced by holistic demand-driven control principles utilizing real-time demand information, such as POS and RFID captured data. The potential effects of reduced lead times and stock levels were of particular importance for perishable products. The project also demonstrated how access to both real-time and historic demand information could be utilized to develop automated decision support and visualization systems supporting supply chain operations.

For global manufacturing operations, important challenges include how to coordinate geographically distributed and diversified networks of plants and facilities and how to control and execute operations according to demand and customer segments. In the Origo project (2006-2008, see www.sintef.no/origo), a Global Control Centre environment was developed to deal with coordination of planning and control tasks in a global network of fishhook production facilities (see Dreyer, Alfnes, Strandhagen, & Thomassen, 2009). A similar project was Optilog (2005-2010, see www.sintef.no/optilog) where the focus was on developing new collaboration models and associated supply contracts that capture the need for improved operational coordination and information sharing along a construction supply chain, from pipe manufacturer, through wholesaler to contractors.

A number of common elements can be seen in the above mentioned examples:

- A rethinking of the traditional manufacturing planning and control (MPC) paradigm, such as MTS, push and batch principles, towards demand-driven control principles such as make-to-order (MTO), pull and continuous, automated replenishment
- Fulfillment of demand based on replenishment and collaboration concepts, instead of traditional ordering and purchasing practices
- Potential for performance improvement and achievement of reduced costs, lead times, stock levels, and throughput time, through sharing of demand information and utilization of updated and real-time information
- Redesign and adjustment of supply chain operations to demand and customers' needs, including coordination and integration of supply chain processes into a unified control model
- Utilization of updated and real-time information in automated decision support systems and control models, enabled by modern ICT solutions

The remainder of this section will describe essential elements of these topics, as well as recent developments and the authors' views on how these developments are expected to shape the future business environment.

The MPC task in a supply chain involves determining what, who, when and how to act, in order to meet customer demand with the exact supply in a coordinated chain (Jonsson & Lindau, 2002; Vollmann, et al., 2005). Each node in the chain cannot be managed in isolation (Shi & Gregory, 1998), and the MPC system must support cross-company processes in a manner that avoids increasing amplifications, stock levels, and lead and response times (Dreyer, et al., 2009). Today, most of the planning and control systems used in supply chain operations are based on the traditions of MTS and MRP/MRP II, where forecasts and expectations of future demand are the main inputs (Zijm, 2000). In addition, economies of scale arguments are frequently used when dimensioning and controlling processes like production, warehousing, distribution and transport. The main planning and control logic of ERP (Enterprise

Resource Planning) systems is still based on aggregation, optimal batch sizes, order quantities, transport frequencies, and sequencing (Alfnes & Strandhagen, 2000). The consequences are that a number of supply chain operations are decoupled from actual end customer demand, and inventories are used as a buffer against uncertainty and fluctuating demand.

The next generation supply chain MPC models are derived from the principle of sharing demand information among all members of the supply chain, so that production, warehousing, distribution and transport can be operated based on demand, rather than expectations and forecasts. Moving the information decoupling point backwards in the supply chain is an essential element in this. The more members of a supply chain that have an undistorted and near real-time view of consumer buying behaviour, the more responsive the supply chain is as a whole (Mason-Jones & Towill, 1999).

In order to develop demand-driven MPC models, several collaborative models for orchestrating supply chain and network activities have been developed. The aim of models such as collaborative planning, forecasting and replenishment (CPFR), vendor managed inventory (VMI) and automated replenishment programs (ARP) is to achieve seamless inter-organizational interfaces by specifying control principles and operations models for the flow of materials and information (e.g. Daugherty, Myers, & Autry, 1999; Mattsson, 2002; Sabath, Autry, & Daugherty, 2001). The main principle is to tie and adjust network operations to customer demand and MTO strategies, instead of the traditional forecast and MTS approaches.

Performance monitoring is essential to ensuring efficient optimisation of operational processes, and, over the last few years, the focus has shifted to incorporate a supply chain perspective (e.g. Chan & Qi, 2003; Holmberg, 2000; Lapide, 2000; Van Hoek, 1998). Supply chain performance measurement requires continuous follow-up, based on a consistent and comparable holistic structure or hierarchy of indicators. These are used to measure the performance of the supply chain as a whole and of the individual members, based on agreed upon strategies, performance targets and priorities. Both reactive or "lagging" indicators showing how the supply chain has performed, and proactive or "leading" indicators showing the efficiency in operational supply chain processes, are required (Bititci, Turner, & Begemann, 2000; Holmberg, 2000; Neely, et al., 2000).

In any supply chain or network, there will be a mixture of principles which must be aligned with the manufacturing environment and the specific control situation of the supply chain. Automation can be built into the control model through the use of ICT tools and a conversion of control principles into mathematical algorithms, expressions and logic (Strandhagen, Alfnes, & Dreyer, 2006), making it possible to replace much of the current manual decision-making with status monitoring, surveillance, and exception handling.

Increasing access to real-time demand and event information in the supply chain is expected to lead to a shift in planning and control concepts towards purer demand- and pull-driven supply chains (Kärkkäinen & Holmström, 2002). Technology advances within areas such as RFID, sensor technology and EPCIS will enable access to more real-time information than the existing technology solutions. RFID tags contain information that can be read from a distance, considerably increasing the number of points where data can be obtained throughout the supply chain, compared with today's barcode system. Combining RFID technology with sensor technology further enhances the intelligence of such data capturing technologies, and will enable the development of intelligent and automated planning and control concepts. Advances in ICT have also enabled the development of control center concepts, where dashboards, real-time monitoring systems and visualization applications support information transparency and control on all process levels

(Boyson, Harrington, & Corsi, 2004; Dreyer, et al., 2009).

A control model can be used as a tool for developing a common understanding and organizing operations (Alfnes, Strandhagen, & Dreyer, 2006). The model is a formalized way of describing the underlying logic and the defined control principles of the MPC, which can be used as a foundation for reengineering and improvement processes. The model consists of six "views" on enterprise operations that should be mapped and modeled: *resources* (machines, equipment, facilities, etc.) *materials*, (products, components and materials), *information* (information elements accessed, stored, processed, and transferred), *processes* (administrative, physical and support processes), *organization* (organizational entities, assignment of responsibilities and authority, etc.), and *control.* Control is the key perspective of the model, describing how operations are organized and controlled in manufacturing and distribution, through the following building blocks (Strandhagen, et al., 2006):

- Control principles and methods; defining main principles for how operations are controlled in the chain and for each operations area (e.g. push, pull and ordering)
- Customer order decoupling points (CODP); dividing the supply chain into parts based on forecasts and parts based on customer orders/demand
- Main operation processes and buffers/stocks
- Operations areas; specifying operations that constitute separate areas of responsibility
- Material flow; specifying the main routes through the operations processes
- Information flow; specifying the flow of information related to the supply chain

Summing up the developments and key issues described above, we see that control models provide a foundation for unified and shared control that enables demand-driven and intelligent control concepts. Such control models should enable flexible networks to respond to customer demand by utilizing their own and suppliers' capacity and resources. However, to avoid conflicting interests, the unified control model should contain a number of carefully designed and collaboratively developed control principles. In addition, access to and sharing of real-time information in the network and the utilization of ICT for information processing, visualization and decision support are key elements in the realization of intelligent and demand-driven supply chain concepts.

TOWARDS A FRAMEWORK FOR INTELLIGENT AND DEMAND-DRIVEN SUPPLY CHAIN CONTROL

The complexity and variety of global supply chains, as described in the previous section, points to the need for capturing a more holistic picture of supply chain planning and control that will support future developments in industrial business cases. In response to this need, a conceptual framework and associated methodology have been developed, based on the above elements and experiences from numerous cases. Capturing control as its key element, the framework is called a c*ontrol model for intelligent and demand-driven supply chains*, consisting of the following parts:

- *A concept;* providing an overview and general description of each of the key elements
- *A set of principles;* describing the shifts required in moving towards demand-driven supply chains based on real-time information
- *Case illustrations;* describing of a set of business cases illustrating application of the concept and associated principles within a number of different industrial sectors,

types of supply chains, product segments and/or particular processes

- *Generic control model;* a generic model template that allows for specific model development for individual business cases
- *Guidelines;* a set of additional and practical guidelines (supporting the general guidelines of the control model methodology) that specifically address issues related to real-time based demand-driven supply chain

At the current state of this research, the concept and the principles have been developed, and the remainder of this chapter will focus on describing their essential elements. A first version of the generic control model has been used as a template in a number of cases and projects where the framework has been applied, and one of these cases is used in this chapter to illustrate the principles and model. The generic model and guidelines are the focus of our ongoing research.

A Concept for Intelligent and Demand-Driven Supply Chains

The concept is the starting point of the control model for intelligent and demand-driven supply chains. Figure 1 shows the individual elements that need to be addressed and interconnected, in order to realize the concept. Each of the elements is described in more detail below.

Control

The control element consists of a description of the processes and principles applied to direct the flow of goods and information in the supply chain. For each process in the supply chain (supply of material, production, assembly, wholesale, retail, etc.) the control principles (including placement of CODP) specify issues such as when, how, and where to act, upon which criteria and in what quantities.

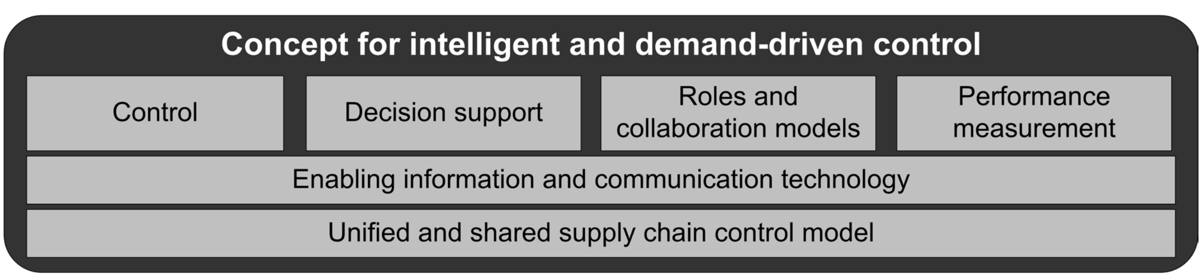
Decision Support

This element is the description of how decision support is applied in supply chain control. It describes all application areas of decision support, with particular focus on applications related to operative decision-making. Examples include the use of optimization methods or advanced simulation tools in evaluation of lot-sizes or order quantities.

Roles and Collaboration Models

Roles and collaboration models outline how supply chain processes are coordinated and who is responsible for what. All processes and information flows that cross organizational boarders, as well as business and organizational mechanisms such as contracts and price instruments, are also described.

Figure 1.



Performance Measurement

Performance indicators show how good and effective supply chain operations are, indicating when and where changes and improvement actions should be made. This description contains the three main application areas of monitoring, cause-effect analysis and management reporting, and it describes the processes of data collection, analysis and presentation, as well as the systems assigned to support such processes.

Enabling ICT

All the other elements of the control model emphasize the need for ICT tools and an architecture which can support transactions, information processing and decision-making in real time. This element of the concept identifies and describes the key technologies and systems applied to support the supply chain control tasks.

Unified and Shared Supply Chain Control Model

All the elements of the control concept are collected into a unified control model. The control model can be seen both as a way to structure and formalize operations activities, and as a fundament for reengineering and improvement processes.

Principles for Intelligent, Demand-Driven Control

Essential to the framework is a set of principles that will support the shift towards demand-driven supply chain control based on real-time information, where each principle is a general rule for how to realize the concept. Table 1 briefly describes the general topic of each principle.

Each of the eight principles should be applied and described in a unified control model. This control model must be established between all partners in the supply chain and communicated to all those involved in supply chain operations. The control model should also be the logical

Table 1. Principles for intelligent, demand-driven control

	Topic	Main principle
1	Capturing and utilizing real-time information	Real-time information on POS data, stock levels, marketing plans, manufacturing and transport status, etc. should be captured and utilized for decision-making.
2	Information sharing	Information should be shared continuously with supply chain partners through databases, information hubs or portals that store, integrate, process and present relevant information from the ICT systems of each actor.
3	Reconsidering the CODP	CODP placement should be differentiated and moved up or down stream in the supply chain according to product and market characteristics.
4	Pull control	Decisions on manufacturing, shipment, etc. should be proactive and based on a “buy one – produce and deliver one” strategy.
5	Supplier-driven replenishment	Replenishment responsibility should be moved upstream in the supply chain (supporting VMI concepts).
6	Individual product and process control	Control of each product or product segment should be determined individually according to product, demand and market characteristics.
7	Automated and integrated decision support	ICT should be used for automated decision-making and provide decision support tools. Human decision-making should be shifted to surveillance and status monitoring.
8	Measuring supply chain performance in real time	A new set of leading and lagging performance measures should be developed and integrated with control parameters. These measures should be implemented in a visual dashboard showing supply chain performance in real time.

basis for the development of supply chain control dashboards or studios. The following sections describe each principle in more detail.

The Principle of Capturing and Utilizing Real-Time Information in the Supply Chain: Information on supply chain operations is often historic data on POS data/demand, stock levels, marketing plans, manufacturing and transport status, etc. and typically based on batch, periodic, or entry point approaches which accumulate demand. Consequently, decisions are made based on the past, not on the current situation. However, technological developments have now enabled access to up-to-date and real-time information, which will lead to a major shift towards supply chain control based on continuous insight into the current situation, thus allowing real-time execution and handling of events. Information on actual events is available, and instead of waiting until the next period or an order-point is passed, the information can immediately be made available to supply chain partners. This is particularly relevant and critical for immediate suppliers of the items in question. Access to real-time information means decisions can be made based on facts and realities rather than on assumptions and expectations, which contributes to a more precise statement of demand, stocking of the right product mix, etc.

The Principle of Information Sharing: The ability to capture real-time information, described in principle (1), provides for improved information quality. However, if information (real-time or otherwise) is not continuously distributed and shared among supply chain partners, the value of information for improved decision-making is highly limited. Since the benefits, in terms of improved supply chain responsiveness, increase as the information order decoupling point is moved up the chain, information on consumer behaviour should be made available as far back in the chain as possible. To increase usefulness of market information, filtering and aggregation routines should be established for each node. The information volume will be significant, and establishment of a unified ICT architecture and development of tools for processing and presenting relevant information elements is necessary. One solution would be the establishment of databases, information hubs and portals to store, integrate, process and present information from the entire supply chain, integrating information from the ICT systems of each supply chain member.

The Principle of Reconsidering the CODP: The access to real-time information will allow reconsideration of the placement of the CODP in general and individually, by product or product group. The move could be upstream, as well as downstream, in the supply chain. Where to place the CODP depends on a number of variables, including product and market characteristics, and a main rule for demand-driven supply chains whether real-time information is available is to move the CODP upstream. An argument for this is that real-time information makes the administrative lead-time near zero, increases responsiveness, and thereby eliminates the need to make or move goods based on forecasts. For products with high value and customer-specific elements, this will be a main principle. For standard products and commodities with long transportation time, where the decision on final destination can be made while the goods are being moved, real-time information can allow moving the CODP as far forward in the supply chain as to the unload process.

The Principle of Pull Control: The overall planning and control principle should be make-to-order according to a "*buy one – produce and deliver one"* strategy. A pull-based solution is therefore essential in demand-driven supply chains. However, with access to real-time information control, decision-making can be shifted from reactive to proactive. A supplier can act and send shipments based on real-time insight into customers' stock levels and thus not have to wait for an order before initiating the appropriate activities. Additionally, since information is made available continuously, the existing control principles based on economy of scale reasoning, such as order point, periodic batch review, economic order quantity (EOQ), etc., will be challenged. Real-time planning and control models differ from more traditional models by:

- Dynamic and continuous determination of order quantities and shipment frequencies
- Individual and automatic product decision-making
- Decision-making based on:
    - A holistic perspective, taking into consideration a number of factors such as sales, stock levels, manufacturing, transport routes, substitution products, packing quantities, etc.
    - Simultaneous handling of information on history, current status and forecasts
    - Aggregated information (from items to product-groups, components and materials)

The most radical shift is that decisions on manufacturing and shipping order quantities and frequencies are dynamic, continuous and event-based. As systems are designed to be triggered by events and automatically recalculate, fixed order quantities and batch processing of information is abandoned.

The Principle of Supplier-Driven Replenishment: In a more traditional control situation, the customer is responsible for communicating a demand signal to the supplier in the form of an order. With access to real-time information, the responsibility for ordering is shifted to the supplier, who replenishes the customer with the exact product quantities at the right time, for instance through VMI solutions. However, such solutions require that the supplier has continuous access to correct information on the customer's demand, stock levels, marketing plans, etc. The need for a more holistic supply chain perspective requires that replenishment decisions (automated or manual) are made as far upstream as possible. The principle is that the upstream supply chain members have the advantage of being able to combine information for all subsequent nodes and can utilize this to adjust capacity and achieve economies of scale benefits.

The Principle of Individual Product and Process Control: Ideally, each product or product group should be dealt with individually, in order to take into account specific product, demand and market characteristics when handling the product through the supply chain. A real-time and demand-driven supply chain should therefore have a clear process orientation, where control is organized around the product and how it flows through the main operations processes (manufacturing, warehousing, distribution, replenishment, etc.). This will typically be in line with concepts such as multi-echelon inventory control, automated replenishment, and collaborative planning, forecasting and replenishment (CPFR). Real-time information, integrated decision support and ICT tools enable such differentiated control models down to an individual product level, thus creating flexibility.

The Principle of Automated and Integrated Decision Support: Real-time information, in combination with a unified control model, can be combined with decision support tools to automate operative decision-making. Such tools include mathematical algorithms, optimisation functionality, ICT, automated analyses, and integrated support tools, such as simulation and what-if analysis. Decisions will not be made in isolation, but rather will be based on a unified foundation, in order to handle trade-offs and alleviate sub-optimization. Through its processing capacity, ICT can identify patterns in demand and supply much more quickly and provide more accurate forecasts for short- and long-term operations. ICT tools can integrate information from several nodes, as well as integrate forecasts, current status and historical data. As the degree of automated decision-making is radically increased, it is essential that information on decisions is presented automatically and visually in dashboards or studio environments for human surveillance and status monitoring.

The of Measuring Supply Chain Performance in Real Time: Within a traditional control paradigm, supply chain operations are evaluated based on past performance, and indicators seldom predict consequences on future performance, thus providing a weak foundation for corrective actions and improvement efforts. With access to real-time information, a new set of performance measures which include leading and lagging parameters should therefore be developed and integrated with control parameters. Real-time key performance indicators (KPIs), showing the exact current performance for the entire supply chain (e.g. lead time, average stock levels, inventory throughput time and fill rate), should be continuously accessible to all partners. Further, ranges for acceptable performance should be established for each KPI. Automatic and electronic procedures can be programmed to generate alerts whenever an indicator falls outside of the agreed ranges. The whole set of performance measures should be integrated in a visual dashboard solution, thereby providing a powerful tool for supply chain performance improvement. Ideally, supply chain members should have access to a flow map showing, in real time, the movement of goods and materials through the supply chain and how the supply chain as a whole is performing.

Case: Control Model in a Fresh Food Supply Chain

In this section, a control model for an intelligent and demand-driven fresh food supply chain will be described in terms of the principles presented above applied to a case study in the Norwegian grocery industry - the production and distribution of processed fresh salads. The information from the case stems from a three-year user-driven R&D project called *Smart Flow of Goods* (2007-2009, see www.sintef.no/smartvareflyt). The project was initiated by major players in the grocery industry (three food manufacturers, two manufacturers of packaging material, two wholesaler – retailer dyads, and two logistics solution providers). The aim of the project was to develop new intelligent and demand driven control models for the supply chain, enabled by RFID technology and intelligent control principles.

The focus and activities in the case were determined in cooperation between practitioners and researchers, based on the specific needs and challenges of the participating organizations and their supply chains. Data was derived using the control model methodology (Alfnes, et al., 2006). Solutions were developed through scientific methods (combined action research and case methodology, use of theory, literature, scientific methods for data collection and analysis, etc.), combined with practical knowledge and experiences from the food and grocery industry.

The following sections provide a description of the fresh food industry and the salad supply chain, including current problems and challenges facing the supply chain members. Then, a new control model will be outlined based on the principles described in the preceding sections.

The Salad Supply Chain

The fresh food industry is influenced by a number of trends affecting most global businesses, e.g. globalization, increasing cost focus, time pressure, environmental concerns, customization, and financial rationalization and consolidation (see e.g. Christopher, 2005; Hofmann & Reiner, 2006; Turban, Leidner, McLean, & Wetherbe, 2006). Food supply chains are complex and form large networks, and the grocery industry's solution to its logistics challenges has tended to set a standard for other industries. The logistics structure is often centrally coordinated, enabling use of cross-docking and terminal facilities in distributing goods to retailers in parallel with direct shipments from producers. Control concepts in the Norwegian grocery industry are dominated by traditional push and forecasting based control. Information on consumer demand is, in many cases, not accessible due to lack of, or infrequent, information sharing, and those who have access to demand information do not necessarily utilize it for control purposes.

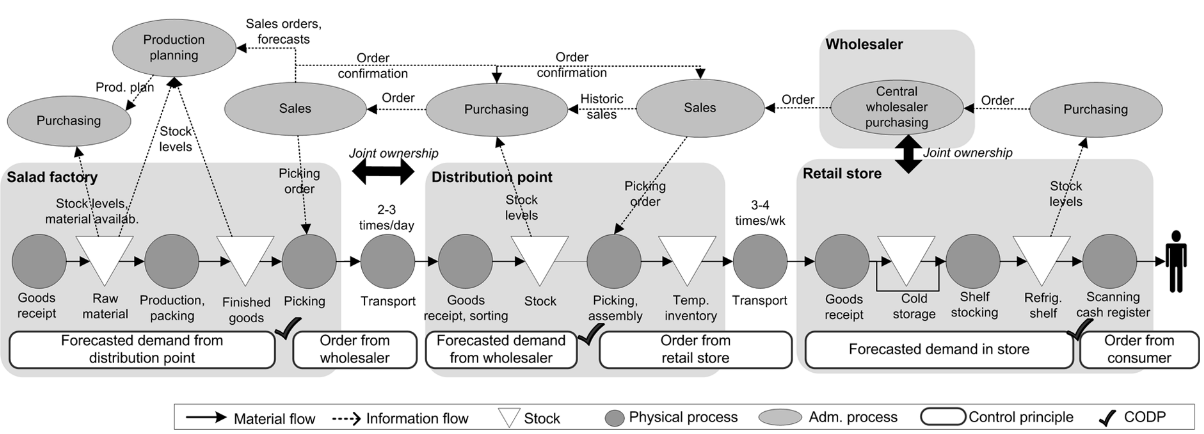
Despite the limited cooperation and information sharing along the food supply chain, Norway is seeing signs of a shift towards more dialogue and collaborative search for industry solutions that can create more efficient and demand-driven supply of food products. The Norwegian grocery market is dominated by a handful of large players within each stage of the supply chain and there is a near full consolidation into four chains of wholesaler – retailer dyads, which control 98% the market. In the past few years, a number of collaborative initiatives have been launched by government and the industry itself, with the aim of increasing efficiency, ensuring supply of safe and high quality food products, reducing waste or scrapping along the food supply chain, etc.

Traditional food items are produced in high volumes with small profit margins. Simultaneously, changing consumer preferences and tough competition are forcing food producers to offer a broader product selection, while continuously reducing costs, which requires increased efficiency in production and logistic operations (Bolseth & Alfnes, 2001).

The case example consists of a supply chain for the distribution of salads from the producer, through a distribution point to a typical grocery store. The focal product consists of various types of processed salads, all with a shelf life of around 8-10 days from the production date. The current supply chain for salad is illustrated in Figure 2, showing supply chain members, material and information flows, stock points, physical and administrative processes, and control principles and associated CODP.

The physical flow starts at the salad factory, where raw materials are supplied from a set of local, national and international primary producers. In total, the factory produces 200 variants, of which salads represent an atypical product, distributed both directly to restaurants and larger institutions and through wholesalers or distribution points to retail stores. The production process for most products consists of quality inspections, cutting, washing, and assembly into various product mixes, before packing and storing. Salads are manufactured to stock. Orders from distributors or wholesalers are used for procurement and production and supply planning. The weekly production plan is based on the principles of optimal batch-sizing. Products are kept in stock

Figure 2.



until orders arrive from distribution points or wholesalers, and turnover for salad products is high, due to the extremely short self life.

From the factory, products are shipped to a distribution point, where retailer orders are assembled. Requirements at the distribution point are calculated by the sales department based on a combination of historic sales data and forecasting techniques. In this particular supply chain, the distribution point is a cross-docking facility, and within a 8 – 10 hour timeframe the products are transported to retail stores. Operations at the distribution point are based on orders from retailers. Accumulated orders are received from a central information system operated by the retail chain's wholesale unit. Orders from each store are placed daily into the central system, based on predefined ordering principles and manual inspection routines. Orders from stores vary in both quantity and items. Throughout the supply chain, products are kept within a strictly regulated temperature zone.

The main challenges and problems associated with current operations and control in the supply chain are:

- A high number of stock points or buffers along the supply chain
- Large amounts of waste/scrapping, due to long lead times and temperature sensitive products
- Forecasting and planning in each node, based on forecasted demand from subsequent node
- Forecasting based on historic sales and extensive manual parameter adjustments, based on experiences and market knowledge
- Limited information sharing and use of available information for operations planning and control
- Limited use of POS data in operations planning and control backwards in the chain

A New Control Model for the Salad Supply Chain

Based on the above challenges and issues, a number of potential improvements to the as is situation were developed, as part of the *Smart Flow of Goods* project. The following section describes how each of the principles above can be addressed through a new control model for the supply chain. Some of the proposed solutions were implemented in the case during the project, while others reflect the ideal intelligent and demand-driven control situation for this supply chain.

The new control model involves changes related to the control and structural aspects of the supply chain. The main principle for the new control model is demand-driven control of the entire supply chain, based on continuous sharing of real-time demand and stock information in retail stores. Also, activities related to order, supply and production should be automated, to as a great an extent as possible, with only exceptions and corrections handled manually.

The new control model is based on real-time POS and inventory information being captured and communicated backwards in the supply chain from retail shops (principle 1). This information is an essential input to the proposed replenishment model between the distributor and the retailer. The information will also be made available to the salad factory, for improved production planning and control. Due to the very short shelf life of salad products, demand information must be made available continuously or daily at worst, to enable the factory and the distribution centre to respond quickly to any changes or disruptions in consumer demand.

Due to the high perishability of processed salads, any time spent in stock reduces the quality of the product. The current finished goods stock point in the factory is mainly there as a demand and transport buffer and therefore represents non-value adding activities and time. The finished goods stock point will therefore be moved downstream, such that only the distribution point has a stock of finished goods. CODP will be at production/packing line in the factory (principle 3), with activities in the supply chain based on a combination of a push and pull strategy (principle 4). Downwards from the factory transport, picking, packing and labeling will be based on the replenishment needs of retail stores. Upwards, production at the factory will be based on a push strategy, where production is executed on forecasted accumulated demand, adjusted with consumer demand information. Production batch sizes will be set according to availability of raw materials and production capacity.

As part of an automated,VMI strategy, the distribution point will assume the responsibility for replenishing all retail stores, based on real-time information on POS and stock levels (principle 5). In this way, stock at the distribution point and the entire retail chain will be controlled according to multi-echelon principles, where the stock levels of the two echelons of the supply chain are considered in concert. Operations at the distribution point will be planned and controlled based on aggregated information about replenishment needs for all retail stores. For retail stores, the VMI solution will eliminate the need for traditional purchasing and order processes, while the manufacturer will be able to control the goods flow, from the purchasing of raw material to production, through access to real-time information on distribution needs.

ICT is a key enabler of the new control model, and the implementation of an efficient infrastructure for capturing and sharing information is essential. A supply chain dashboard will be implemented to store, aggregate and visualize demand information (principle 7). In addition, decision support tools will be integrated into the dashboard. POS data and stock information from retail stores will be transferred to the system on a continuous or daily basis, where the distributor and factory can use the information in their own planning.

The new IT infrastructure will also involve the acquisition of commercially available software to support the VMI solution. Such software will calculate the replenishment needs of finished goods for the distribution point and all retail stores, based on historic demand patterns, lead time, demand uncertainty, specified service levels, and information on ongoing and upcoming campaigns and other market activities.

A jointly developed system of performance indicators will be integrated into the supply chain dashboard (principle 8). These parameters will enable supply chain members to continuously measure and monitor performance against established objectives for the supply chain and each individual member. The indicators will form a hierarchy showing how the performance of individual supply chain members affects overall supply chain performance. Lagging indicators include sales, productivity, quality control costs, and rush order costs, while leading indicators include measures of service levels, stock levels, stock values, returns, product and delivery quality, lead times, and supply chain responsiveness.

This proposed control model refers only to one product group: processed salads. Since this is an atypical product for the factory and the rest of the supply chain, separate control models should be created for other products and product groups. Although there will be differences between the specific models, they will all be based on the same principles of demand-driven control enabled by sharing of real-time information. Each model should take into consideration the characteristics of the products, their demand patterns and variability, particular requirements related to production and distribution processes, etc.

Issues, Controversies, Problems

The previous sections of this chapter have highlighted possibilities and the essential elements managers must deal with in order to realize intelligent and demand-driven supply chain operations.

However, the control model presented is based on a normative approach for dealing with complexity and improving supply chain efficiency. The model was developed using a general approach based on some simplifications and introducing several elements and principles as building blocks in a new environment for operating supply chains. In addition to the recommendations and principles set forward in this chapter, there are a number of other issues which supply chain managers will have to address, both within their own organizations and in their supply chains, before such a normative model can be applied in a real supply chain setting. Some of these issues, controversies and problems are discussed below.

Efficient Coordination Across Organizational and Geographical Boundaries: Although the case dealt with in this chapter focuses on a domestic supply chain, the issues discussed and the model developed is possibly even more relevant in the context of global network operations. In any situation where the scope of operations span is beyond a single organization single-site, the planning and control task takes on a supply chain or network perspective. The high complexity and coordination challenges involved in supply chain operations therefore require a holistic and integrated approach to planning and control, where one single, agreed-upon control model should support seamlessly integrated processes across organizational and geographical boundaries. However, the unified supply chain control model presented here does not guarantee efficient coordination of all operations, particularly since it does not solve the issues related to the lack of a central coordinating facility or unit in a supply chain of autonomous members. The joint development of a unified control model is, however, a step in the right direction.

Dealing with Inter-Organizational Issues Related to Information Sharing, Diverging Interests, and the Power and Trust Balance: The power position and relative strength of supply chain members is discussed both in the literature and practice. In the food supply chain, many retailers have gained a major strategic position through integration of retailing and wholesaling activities, allowing them to control access to demand information. This information is frequently not shared with other partners in the supply chain, because retailers are worried that suppliers might use it to gain a competitive advantage. In the normative framework explored in this chapter it is assumed that retailers share demand information, even though, in reality, retailers tend to be restrictive in sharing this information. Also, in many instances, suppliers lack models for how to utilize demand information to improve their own operations.

Ability to Handle Dynamics, Uncertainty and Variation: For supply chain operations, it is extremely challenging and resource-demanding to cope with variation and uncertainty. This means that the models suggested in this chapter need to be dynamic and flexible, which somewhat breaks with the normative approach of the models. Still, this is one of the most important issues to address, in order to handle supply chain complexity.

Simultaneous Participation in Multiple Supply Chains: Each partner in a supply chain needs to deal with the fact that they most often are part of several supply chains simultaneously. Thus, solutions developed in cooperation with partners in one supply chain might not be appropriate for all or indeed any of the other supply chains the company participates in. Also, solution working well for one partner in a particular supply chain may not be successful or applicable for other partners due to the heterogeneous nature of business processes and relationships. The "one size fits all" approach is hardly applicable in a supply chain setting, thus, a major challenge is categorizing and finding principles which simultaneously differentiate and harmonize.

Use of and Investment in ICT: Electronic data capture and information processing and sharing requires considerable investment in technology (RFID, dashboards, EPCIS, etc.) and competence. This can be particularly challenging for small and medium sized enterprises (SME) with limited resources. For many companies and sectors, this is a contributing factor to the limited interest in and application of new technologies, such as RFID and automated replenishment systems. The investment level, the upgrading of competence and the pace of technological development are all obstacles to a higher degree of ICT application.

Standardization of Information and Interfaces: Information has to flow freely with low friction and be understandable for all members of a supply chain, in order to create value. This means that data should be standardized and that the organizational and technology interfaces should be harmonized. Thus, advances within global information standards are highly important in enabling a greater degree of information sharing in a supply chain.

Explosion in Information volume and Information Security: Developments in data capturing technology, mobile and wireless communication platforms, track and trace solutions, etc. have led to an explosion in information volume. The ability and capacity to deal with ever increasing information volumes in an appropriate and secure manner is an issue of high relevance. This also includes the question of data protection in general, and personal information in particular.

Solutions and Recommendations

Despite the challenges and issues outlined above, there are developments and initiatives which can provide support in overcoming some of the obstacles. This section outlines some recommendations and solutions which can assist supply chains in realizing the concept of intelligent and demand-driven supply chains.

At present, there is a lack of business cases which demonstrate solutions using the proposed control model. By implementing a well-defined model in a variety of representative contexts and sectors, the potential effects and implications for supply chains can be evidenced. One first step could be to implement selected elements of the control model. Although the technical solutions for capturing real-time information are still not in place in most supply chains, each supply chain member already captures large amounts of data which can be very beneficial to their supply chain partners. A manufacturer might find that regular updates containing the POS data, stock levels and marketing plans of key customers could be used to improve production planning and control.

Development within the area of system standards is continuously evolving. Through, for instance, the EPCIS standards and interfaces, the ability to exchange information with supply chain partners is improving. Further, technology and system developments with regards to integration, functionality and cost are paving the way for smoother and more efficient supply chain cooperation. Data collection devices such as RFID, antennas, mobile broadband solutions, etc. will soon be reaching a level of functionality and cost that make implementation on a broader scale realistic.

Of particular importance in the realization of efficient supply chain collaboration and information sharing is the development and implementation of supply chain control dashboards. Such dashboards originate from executive information systems and control rooms in the process industry. A dashboard integrates relevant information from all supply chain partners' information systems (e.g. enterprise resource planning, supply chain management, advanced planning, manufacturing execution, etc.), thus supporting a true supply chain perspective, enabling visualization of performance indicators, and supporting operations, planning and control on all process levels (Dreyer, et al., 2009). In the Automed R&D project referred to earlier, a prototype of a dashboard for pharmacy products was developed in cooperation with Oracle. *Figure 3* shows a screen shot of the "Stock availability" function, where forecasts and sales for a stock keeping unit (SKU) is shown graphically for the manufacturer, supplier, wholesaler, and, in this case, aggregated for all pharmacies

Figure 3.

in the pharmacy chain. The dashboard allows aggregation and drill-down between levels, to get more detailed stock level information from an overall supply chain status and down to individual pharmacies.

The use of supply chain dashboards can either be individually by employees within each company, or as part of a collaborative environment where groups and teams are involved in the execution of supply chain activities. In a studio environment (virtual or physical) members of distributed and complex supply chains can cooperate and make decisions in a holistic and integrated manner. Distributed and linked studios in a supply chain can serve as information and communication nodes that enable integration of operations activities and distributed and interactive supply chain planning, across geographical distances and organizational levels. Supply chain studios provide members with access to network-wide real-time information (Wang & Wei, 2007), enable visualization of the available information (Boyson, Corsi, & Verbraeck, 2003) through, for instance, supply chain dashboards, secure interaction between advanced ICT based decision support tools and human decision-making (Barthélemy, Bisdorff, & Coppin, 2002), and create a coordinated and collaborative environment (Deek, Tommarello, & McHugh, 2003) for planning, control and decision-making.

FUTURE RESEARCH DIRECTIONS

True real-time operations of supply chains are still seen only very scarcely and only at very limited subparts of supply chains, for instance, between two members of a complete chain. There are numerous development and research areas that need to be addressed, in order to realize real-time controlled supply chains.

Within the operations management area, the following issues remain to be addressed:

- Further development of control principles and collaboration mechanisms, from the level of principles into guidelines, rules, formulas and algorithms.
- Shift toward applying optimization and simulation (operations research) at a truly operative level, integrating optimization and simulation with control algorithms and systems like ERP, DRP (distribution requirements planning), scheduling, etc.
- New perspective and role of performance measurement, shifting perspective from a focus on lagging indicators to support for on-line decision-making
- Study and understand obstacles and possible solutions with regards to power issues, willingness to cooperate, collaboration models, incentive mechanisms, etc. in different industrial sectors.
- Study and understand obstacles and possible solutions regarding the human role and behavior in the new paradigm, where human and manual decision-making is replaced by surveillance, status monitoring and automated decision support through visual monitoring facilities.

Regarding new technology, including the ICT perspective, relevant and important issues for future R&D include:

- General development of ICT performance: processing capacity versus cost, system integration capability, visualization facilities, man-machine interfaces, etc.
- Future of RFID, sensor and mobile technology for close-to-zero cost of tracking individual items
- Standardization of data exchanges based on global coding (EPC)

It will be in the true interplay among these R&D challenges that we will see the realization of future demand-driven supply chains.

CONCLUSION

Increasing the competitiveness of supply chain systems requires a supply chain which is integrated and coordinated, where operations are driven by demand and real-time principles. A supply chain is, by nature, a complex and heterogeneous system with a wide range of products, processes and companies, as well as diversified material and informational flow structures. A unified planning and control model is an efficient approach to coping with the associated issues and challenges. In this chapter, we have explored and developed such an approach in the form of a framework for intelligent and demand-driven supply chain control, where we assume that all relevant information is made available to all partners in real time. In order to capture a more holistic picture of supply chain planning and control that supports future developments in industrial business cases, the framework consists of a concept and a number of associated principles. Based on the control model methodology, these principles have been demonstrated through a control model for processed salad products.

The following is a brief description of the main elements of the principles. Replenishment and shipping decisions should be based on real-time information about POS data, stock levels, marketing plans, manufacturing and transport status, etc. CODP placement should be differentiated and moved upstream or downstream in the supply chain, according to product and market characteristics. Manufacturing and shipment of goods should be based on the principle of *"buy one – produce and deliver one"*. Replenishment responsibility should be moved upstream in the supply chain. Human and manual decision-making should be replaced by surveillance, status monitoring and automated decision support through visual monitoring facilities. The control of each product or product segment should be determined individually according to product, demand and market characteristics. Information hubs should be established to collect, store and present information from the entire supply chain, integrating information from the ICT systems of each supply chain member. A new set of performance measures, showing leading and lagging parameters, should be developed and integrated with control parameters. Finally, these measures should be implemented in a visual dashboard and made available for monitoring.

Implementation of truly demand-driven supply chains is dependent on developments and improvements within technology and standards. Solutions that are affordable for ordinary companies and SMEs are expected to become readily available in the near future. A major implementation obstacle, however, is the lack of trust, open sources and collaboration between supply chain members – issues that require further research. The concept and principles proposed in the chapter have already proven their potential through a number of cases, where results have been achieved, and further changes and solutions are being developed.

REFERENCES

Alfnes, E. (2005). *Enterprise Reengineering - A strategic framework and methodology*. Trondheim, Norway: Doctoral Thesis, Norwegian University of Science and Technology.

Alfnes, E., & Strandhagen, J. O. (2000). Enterprise Design for Mass Customization - the Control Model Methodology. *International Journal of Logistics*, *3*(2).

Alfnes, E., Strandhagen, J. O., & Dreyer, H. C. (2006). An Enterprise Mapping Tool: The Operations Model. In K. Mendibil & A. Shamsuddin (Eds.), *13th International EurOMA Conference: Moving up the value chain*. Glasgow, UK: University of Strathclyde.

Barthélemy, J. P., Bisdorff, R., & Coppin, G. (2002). Human centered processes and decision support systems. *European Journal of Operational Research*, *136*(2), 233–252. doi:10.1016/S0377-2217(01)00112-6

Bititci, U. S., Turner, T., & Begemann, C. (2000). Dynamics of Performance Measurement Systems. *International Journal of Operations & Production Management*, *20*(6), 692–704. doi:10.1108/01443570010321676

Bjartnes, R., Strandhagen, J. O., Dreyer, H. C., & Solem, K. (2008). Intelligent and Demand Driven Manufacturing Network Control Concepts. In JOMSA (Ed.), *The Third World Conference on Production and Operations Management*. Tokyo, Japan.

Bolseth, S., & Alfnes, E. (2001). How to Achieve Agility in Food and Drink Manufacturing. In COMA'01 (Ed.), *International Conference on Competitive Manufacturing (COMA'01)*. Stellenbosch, South Africa.

Bowersox, D. J., Closs, D. J., & Stank, T. P. (1999). 21st Century Logistics: Making Supply Chain Integration a Reality. *Council of Logistics Management*.

Boyson, S., Corsi, T., & Verbraeck, A. (2003). The e-supply chain portal: a core business model. *Transportation Research Part E, Logistics and Transportation Review*, *39*(2), 175–192. doi:10.1016/S1366-5545(02)00046-7

Boyson, S., Harrington, L. H., & Corsi, T. (2004). *In Real Time: Managing the new Supply Chain*. Westport, CT: Praeger.

Chan, F. T. S., & Qi, H. J. (2003). Feasibility of Performance Measurement System for Supply Chain: A process-based approach and measures. *Integrated Manufacturing Systems*, *14*(3), 179–190. doi:10.1108/09576060310463145

Christopher, M. (1998). *Logistics and supply chain management: strategies for reducing costs and improving service* (2nd ed.). London: Financial Times/Prentice Hall.

Christopher, M. (2005). *Logistics and Supply Chain Management: Creating Value Adding Networks*.London: FT Prentice Hall.

Clark, T. H., & Lee, H. G. (2000). Performance, interdependence and coordination in business-to-business electronic commerce and supply chain management. *Information Technology Management*, *1*(1), 85–105. doi:10.1023/A:1019108621684

Daugherty, P. J., Myers, M. B., & Autry, C. W. (1999). Automatic replenishment programs: an empirical examination. *Journal of Business Logistics*, *20*(2), 63–82.

Deek, F. P., Tommarello, J. D., & McHugh, J. A. (2003). A model for collaborative technologies in manufacturing. *International Journal of Computer Integrated Manufacturing*, *16*(4-5), 357–371. doi:10.1080/0951192031000089246

Dreyer, H. C., Alfnes, E., Strandhagen, J. O., & Thomassen, M. K. (2009). Global supply chain control systems: a conceptual framework for the global control centre. *Production Planning and Control*, *20*(2), 147–157. doi:10.1080/09537280802705013

Ellinger, A. E., Taylor, J. C., & Daugherty, P. J. (1999). Automatic Replenishment Programs and Level of Involvement: Performance Implications. *International Journal of Logistics Management*, *10*(1), 25–36. doi:10.1108/09574099910805914

Hofmann, P., & Reiner, G. (2006). Drivers for improving supply chain performance: an empirical study. *International Journal of Integrated Supply Management*, *2*(3), 214–230. doi:10.1504/IJISM.2006.008594

Holmberg, S. (2000). A system perspective on supply chain measurement. *International Journal of Physical Distribution & Logistics Management*, *30*(10), 847–868. doi:10.1108/09600030010351246

Jonsson, P., & Lindau, R. (2002). The supply chain planning studio: utilising the synergic power of teams and information visibility. In Solem, O. (Ed.), *NOFOMA annual conference for Nordic researchers in logistics: Promoting logistics competence in industry and research* (pp. 115–130). Trondheim, Norway: Norwegian University of Science and Technology.

Kärkkäinen, M., & Holmström, J. (2002). Wirless product identification: enabler for handling efficiency, customisation and information sharing. *International Journal of Supply Chain Management 7, 4,* 242-252.

Lapide, L. (2000). True Measures of Supply Chain Performance. *Supply Chain Management Review, July/August*, 25-28.

Mason-Jones, R., & Towill, D. R. (1999). Using the Information Decoupling Point to Improve Supply Chain Performance. *The International Journal of Logistics Management*, *10*(2), 13–26. doi:10.1108/09574099910805969

Mattsson, S. A. (2002). *Logistik i försörjningskedjor*. Lund: Studentlitteratur.

Min, S., & Mentzer, J. T. (2004). Developing and Measuring Supply Chain Management Concepts. *Journal of Business Logistics*, *25*(1).

Neely, A., Mills, J., Gregory, M., Richards, H., Platts, K., & Bourne, M. (2000). Performance Measurement System Design: Developing and Testing a Process-Based Approach. *International Journal of Operations & Production Management*, *20*(10), 1119–1145. doi:10.1108/01443570010343708

Reichart, A., & Holweg, M. (2006). *Research Methods in Supply Chain Management: a critical Review*. Paper presented at the Euroma conference.

Sabath, R. E., Autry, C. W., & Daugherty, P. J. (2001). Automatic replenishment programs: The impact of organizational structure. *Journal of Business Logistics*, *22*(1), 91–105.

Shi, Y., & Gregory, M. (1998). International manufacturing networks - to develop global competitive capabilities. *Journal of Operations Management, 16*(2, 3), 195-214.

Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2008). *Designing and Managing the Supply Chain. Concepts, Strategies, and Case Studies* (3rd ed.). Boston: McGraw-Hill/Irwin.

Slack, N., Chambers, S., & Johnston, R. (2001). *Operations Management* (3 ed.). Washington, DC: Financial Times, Prentice Hall.

Strandhagen, O., Alfnes, E., & Dreyer, H. C. (2006). Supply Chain Control Dashboards. In E. Bendoly (Ed.), *17th Annual Conference of the Production and Operations Management Society (POMS): "OM in the New World Uncertainties"*. Boston.

Turban, E., Leidner, D., McLean, E., & Wetherbe, J. (2006). *Information Technology for Management - transforming organizations in the digital economy*. Hoboken, NJ: John Wiley & Sons inc.

Tyndall, G., Gopal, C., Partsch, W., & Kamauff, J. (1998). *Supercharging supply chains: New ways to increase value through global operational excellence*. New York: John Wiley & Sons.

Van Hoek, R. I. (1998). Measuring the Unmeasurable - Measuring and Improving Performance in the Supply Chain. *Supply Chain Management*, *3*(4), 187–192. doi:10.1108/13598549810244232

Vollmann, T. E., Berry, W. L., Whybark, D. C., & Jacobs, R. F. (2005). *Manufacturing Planning and Control for Supply Chain Management*. Boston: McGraw-Hill.

Wang, E. T. G., & Wei, H. L. (2007). Interorganizational governance value creation: Coordinating for information visibility and flexibility in supply chains. *Decision Sciences*, *38*, 647–674. doi:10.1111/j.1540-5915.2007.00173.x

Zijm, W. H. M. (2000). Towards Intelligent Manufacturing Planning and Control Systems. *OR-Spektrum*, *22*, 313–345. doi:10.1007/s002919900032

ADDITIONAL READING

Bottani, E., & Rizzi, A. (2008). Economical assessment of the impact of RFID technology and EPC system on the fast-moving consumer goods supply chain. *International Journal of Production Economics*, *112*, 548–569. doi:10.1016/j.ijpe.2007.05.007

Boyson, S., Harrington, L. H., & Corsi, T. (2004). *In Real Time: Managing the new Supply Chain*. Westport, CT: Praeger.

Chopra, S., & Meindl, P. (2007). *Supply Chain Management: Strategy, planning and operations*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review*, *75*(2), 105–116.

Holmqvist, M., & Stefansson, G. (2006). Smart Goods And Mobile RFID: A case with innovation from Volvo *Journal of Business Logistics, 27*(2), 251-272.

Ross, A. D., Twede, D., Clarke, R. H., & Ryan, M. (2009). A framework for developing implementation strategies for a radio frequency identification (RFID) system in a distribution center environment. *Journal of Business Logistics*, *30*(1), 157–184.

KEY TERMS AND DEFINITIONS

Control Dashboard: A decision support tool which presents and visualizes key performance measurements that are used to support control processes in a company or supply chain/network.

Control Model: A formalized description of how operations are organized and controlled in manufacturing and distribution, including control principles and methods, CODP, main operation processes, buffers and responsibility areas, and material and information flow.

Control Principle: A definition of what initiates processes in the system. In a pull system the amount of work in process is limited by orders from the subsequent process, while a push system has no limits and work is initiated based on forecasts.

Control Studio/Centre: A physical or virtual workspace and interface supporting collaboration between supply chain planning teams and local managers at different nodes of a network/supply chain. Information to support decision-making is made available through modern and integrated ICT solutions.

Customer Order Decoupling Point (CODP): The point in the supply chain where demand changes from dependent to independent, where the firm becomes responsible for determining the timing and quantity of material to be purchased, made, or finished (Vollmann, et al., 2005); usually coinciding with an important stock point from which customers are supplied.

Information Decoupling Point: The point in the information pipeline where market driven and forecast driven information flows meet, i.e. where information on real customer demand is available to upstream supply chain participants.

Point of Sale (POS) Data: Information about an individual sales transaction (product code, price, number of units), traditionally captured by a bar code scanner at the cash point of a retail outlet. The information is stored in a database and often consolidated on chain level.

Supply Chain Planning and Control: The task of determining what, who, when and how to act to efficiently manage the flow of material and information in order to meet customer demand with the exact supply in a coordinated chain.

ENDNOTE

1 The research presented in this chapter has been financed by the Research Council of Norway and participating industrial companies.

This work was previously published in Managing Global Supply Chain Relationships: Operations, Strategies and Practices, edited by Barbara Flynn, Michiya Morita and Jose Machuca, pp. 49-70, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 18
Reducing Design Margins by Adaptive Compensation for Thermal and Aging Variations

Zhenyu Qi
University of Virginia, USA

Yan Zhang
University of Virginia, USA

Mircea Stan
University of Virginia, USA

ABSTRACT

Corner-based design and verification are based on worst-case analysis, thus introducing over-pessimism and large area and power overhead and leading to unnecessary energy consumption. Typical case-based design and verification maximize energy efficiency through design margins reduction and adaptive computation, thus helping achieve sustainable computing. Dynamically adapting to manufacturing, environmental, and usage variations is the key to shaving unnecessary design margins, which requires on-chip modules that can sense and configure design parameters both globally and locally to maximize computation efficiency, and maintain this efficiency over the lifetime of the system. This chapter presents an adaptive threshold compensation scheme using a transimpedance amplifier and adaptive body biasing to overcome the effects of temperature variation, reliability degradation, and process variation. The effectiveness and versatility of the scheme are demonstrated with two example applications, one as a temperature aware design to maintain I_{ON} to I_{OFF} current ratio, the other as a reliability sensor for NBTI (Negative Bias Temperature Instability).

INTRODUCTION

IT infrastructure has become an essential foundation of human society. Information storage and computation span every aspect of our daily life, from personal entertainment and education, to government and business operation. Moreover, for real-time service providers, data centers and their connections to outside terminals need to be running for a long time without interruption. Examples are banking systems and servers, online shopping and tracking, supply chain management, and multimedia streaming. The unprecedented increasing dependence and requirement on transaction avail-

DOI: 10.4018/978-1-4666-1945-6.ch018

ability and speed call for the deployment of more high performance and reliable computing devices.

However, this leads to at least two problems. At the macroscale level, without careful design and planning the environmental impact due to technology development is unsustainable. For example, the energy cost of IT infrastructure and its carbon footprint doubles in less than five years (Pedram, 2009). Technology-equipped buildings occupy 70% of total electricity usage and emit 40% of greenhouse gases annually in the US (Kleissl & Agarwal, 2010). At the microscale level, high performance and reliability could be at odds for concurrent circuit designs based on the semiconductor technology. For example, high performance designs usually come with high power/temperatures and high stress levels on devices, leading to thermal and reliability problems. In particular, they can cause the instability (including both short time variation and long time shifting) of the threshold voltage of the basic circuit element—CMOS transistors.

The threshold voltage, or V_{TH}, is a critical parameter in circuit design. V_{TH} determines the current flowing through the transistor at least linearly, and in turn affects its equivalent resistance, transconductance, pull-up or pull-down strength and ultimately the speed and power of circuits. Even in advanced sub-micron technologies the dependence of the transistor current on V_{TH} can still be quite super-linear when all short channel effects are considered (Qi & Stan, 2007). The effects in the sub-threshold region are even more prominent since the leakage current grows exponentially as V_{TH} decreases.

Although threshold voltages are specified by technology vendors and treated as fixed and static values in simulations at design time, run-time factors like temperature variability and reliability degradation, as well as process variation, can considerably change V_{TH} both in the short and long term for individual transistors. With the small margins of most circuit designs today, even a relatively small amount of shift can have a significant impact or even lead to circuit failure. On the other hand, preventing these failures by initial overdesigning with large upfront built-in margins can impose large penalties in area and power. For example, an optimized reliability-aware sizing algorithm proposed in (Paul, Kang, Kufluoglu, Ashraful Alam, & Roy, 2006) still incurs an area overhead from 3.31%-13.6% for ISCAS testbench circuits. This conventional method variation by overdesign to tolerate reliability hazards, thermal issues and process leads to unsustainable increase of cost both for the designer and the users.

In this chapter, we propose an adaptive threshold compensation scheme using a transimpedance amplifier and adaptive body biasing to compensate the V_{TH} shift dynamically. The effect of body biasing on modulating transistor threshold voltages has been heavily exploited in recent years to achieve low power and combat process variation (Karnik, Borkar, & De, 2002; Tschanz, et al., 2002). The scheme employed in this chapter works for a range of applications including, but not limited to, manufacturing process variations. Two example applications are presented in this chapter. The first compensates for temperature induced I_{ON} to I_{OFF} mismatch and the second compensates for threshold shifting due to the long time aging/reliability effect.

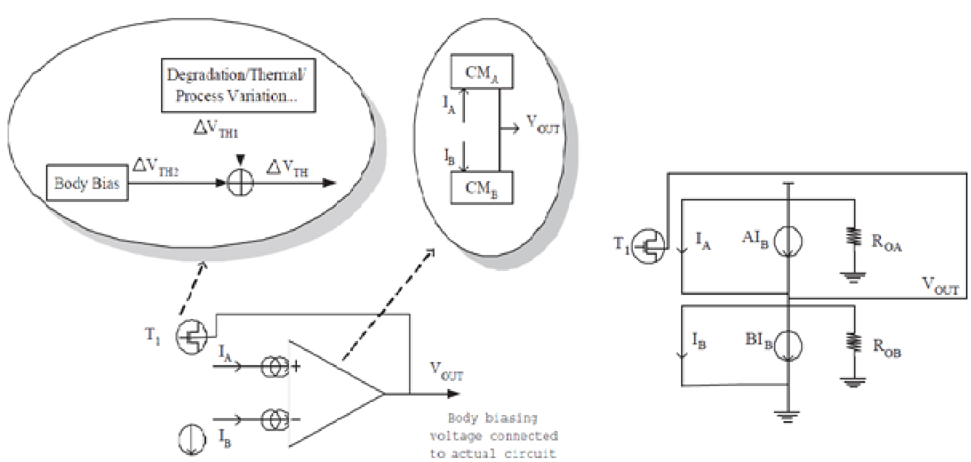
Adaptive compensation of threshold voltage reduces necessary design margins, and enables typical case-based design instead of the traditional worst case-based (or corner based) design. Worst case-based design leads to over-pessimism and unnecessarily large margins/guardbands for the majority of components that do not fall on the tail of the distribution caused by manufacturing or runtime variations. The over-designed components not only take chip real estate and impose non-recurring penalty for designers and manufacturers, but also consume extra power and impose a recurring penalty on the consumer and the environment. The proposed module can also be deployed in large numbers and distributed across the whole chip for adaptive compensation at a finer granularity, and thus achieve more aggressive savings. We believe distribution and adaptation are two key strategies for low-power energy-conserving sustainable computing, and both are embodied in the proposed threshold voltage compensation scheme.

ADAPTIVE THRESHOLD COMPENSATION SCHEME

Figure 1 shows the proposed threshold compensation scheme. It is mainly composed of two parts, the first being the transistor, T1, which could either be a PMOS or an NMOS. T1 serves as a 'proxy' for the transistors in the actual circuit whose threshold voltages need to be compensated due to various reasons including, but not limited to, process and temperature variation and reliability degradation due to aging. The net effect of these factors is the threshold voltage shift noted as ΔV_{TH1} in Figure 1. Ideally, we want to maintain the original V_{TH} value of T1 by offsetting ΔV_{TH1} by ΔV_{TH2}, a voltage shift in the opposite direction of ΔV_{TH1}. In this scheme, ΔV_{TH2} comes from the body biasing effect of T1. Providing the appropriate body bias voltage V_{OUT} to generate ΔV_{TH2} is the goal of the second part of the compensation circuit, shown as a differential transimpedance amplifier on the right.

The transimpedance amplifier mainly comprises two sets of current mirrors, denoted in Figure 1 as CM_A and CM_B. The two differential inputs are the current I_{DS} from the proxy transistor T1, and a reference current I_B. I_A and I_B are fed into CM_A and CM_B separately, and propagated along to the last stage after being amplified with current gains A and B. Notice that the arrows on I_A and I_B simply mean that they are input variables of the circuit, not necessarily the direction of the currents. The composite current gains A and B can be achieved with multiple stages of current mirrors within CM_A and CM_B. The last stage settles to an equilibrium state only if $A \times I_{ON} = B \times I_{OFF}$. This balance provides the output voltage V_{OUT}. V_{OUT} is the body biasing voltage connected to the actual circuit component that needs threshold compensation. It is also fed back to the bulk of the proxy transistor T1 to complete the closed loop system.

Figure 1. The proposed transimpedance amplifier as a closed-loop feedback system and its equivalent small signal circuit model



Threshold Compensation Capability Analysis

The threshold compensation system can be analyzed with the equivalent small signal circuit also shown in Figure 1. Since later we will demonstrate an example application of the threshold compensation circuit for NBTI, the following analysis is performed assuming the proxy transistor T1 is a PMOS, since NBTI only affects PMOS transistors. For the ease of notation, all V_{TH} and ΔV_{TH} carry signs within themselves. Therefore, for a PMOS transistor in the saturation mode, according to the alpha-power law model (Sakurai & Newton, 1991).

$$I_A = K\frac{W}{L}(V_{DD} + V_{TH})^{\alpha} \tag{1}$$

Notice that for PMOS transistors V_{TH} is negative. The meanings of the parameters are summarized in Table 1.

Applying Taylor expansion:

$$\Delta I_A = \alpha K\frac{W}{L}(V_{DD} + V_{TH})^{\alpha}\frac{\Delta V_{TH}}{(V_{DD} + V_{TH})} \tag{2}$$

Referring to Figure 1, here:

$$\Delta V_{TH} = \Delta V_{TH1} + \Delta V_{TH2}$$

The output body bias voltage:

$$V_{OUT} = (AI_A - BI_B)\frac{R_{oA}R_{oB}}{R_{oA} + R_{oB}} \tag{3}$$

R_{OA} and R_{OB} are the small signal output resistance of the two current controlled current sources AI_A and BI_B respectively. Assuming the reference, current I_B stays constant:

Table 1. Parameters used in the derivation

DEFEINITION/DESCRIPTION OF PARAMETER	
PARAMETERS	DESCRIPTION
W, L	width and length of proxy transistor T1
K	device-transconductance parameter of T1
α	short channel coefficient of T1
V_{DD}	power supply voltage for T1
V_{TH}	actual threshold voltage of T1, signs embedded
V_{TH0}	threshold voltage of T1 without body effect
ΔV_{TH1}	threshold voltage shift of T1 due to non-ideal factors
ΔV_{TH2}	threshold voltage change of T1 from body biasing
ΔV_{TH}	net V_{TH} change, $\Delta V_{TH} = \Delta V_{TH1} + \Delta V_{TH2}$
	body effect coefficient, or bulk threshold parameter of T1
I_A	current from proxy transistor T1, and input to CM_A
I_B	reference current, and input to CM_B
A,B	current gain of CM_A and CM_B
R_{OA}, R_{OB}	small sig. output resistance of the last stage
V_{OUT}	output body biasing voltage to actual circuits

$$\Delta V_{OUT} = A\frac{R_{oA}R_{oB}}{R_{oA} + R_{oB}}\Delta I_A \tag{4}$$

Equations (2) and (4) can also be obtained by taking derivatives directly. Finally, considering the body biasing effect, the threshold voltage of T1 is (Sakurai & Newton, 1991):

$$V_{TH} = V_{TH0} + \gamma(\sqrt{2\varphi_F - V_{BS}} - \sqrt{2\varphi_F}) \tag{5}$$

This can be further simplified (Sakurai & Newton, 1991) to:

$$V_{TH} = V_{TH0} - \gamma_1 V_{BS} \tag{6}$$

thus:

$$\Delta V_{TH} = -\gamma_1 \Delta V_{BS} \tag{7}$$

Notice that this ΔV_{TH} is essentially the ΔV_{TH2} in Figure 1. Also for PMOS, ΔV_{TH} and V_{BS} have embedded negative and positive signs.

Combining Equations (2), (4) and (7), we get:

$$\begin{aligned} &\Delta V_{TH2} \rightarrow = \\ &\alpha K \frac{W}{L}(V_{DD} + V_{TH})\alpha \frac{1}{(V_{DD} + V_{TH})} \\ &\gamma A \frac{R_{oA} RoB}{R_{oA} + R_{oB}} (\Delta V_{TH1} + \Delta V_{TH2}) \\ &\quad = C(\Delta V_{TH1} + \Delta V_{TH2}) \end{aligned} \tag{8}$$

where:

$$C = -aK \frac{W}{L} \frac{(V_{DD} + V_{TH})^{\alpha}}{(V_{DD} + V_{TH})} \gamma_1 A \frac{R_{oA} R_{oB}}{R_{oA} + R_{oB}}$$

and $C>0$.

Now we rearrange Equation (8):

$$\frac{\Delta V_{TH2}}{\Delta V_{TH1}} = -\frac{C}{1+C} \tag{9}$$

Equation (9) is quite interesting. Remember that the ultimate goal of the whole system is to use ΔV_{TH2} to compensate the threshold voltage change ΔV_{TH1} caused by non-ideal factors, therefore ideally we would like $\Delta V_{TH2} = -\Delta V_{TH1}$. However, Equation (9) demonstrates that the exact equality is really impossible. However, the larger the constant C is, the more compensation we achieve and the closer the two values can be to each other—C can be viewed as the open loop gain of the feedback system. Increasing the proxy transistor size, the current mirror gain and the last stage output resistance are all possible ways to maximize C. While this opens up optimization opportunities for power and area, these 'knobs' are *not* independent of each other, and the optimization *cannot* be carried out separately, due to the analog nature of the circuit. An important consideration is that the design should always guarantee the correct working range of the current mirrors. How to maximize the gain and quantify the tradeoff of parameters given an area or power budget is our ongoing work.

Finally, notice that although the above analysis assumes a constant reference current I_B, a similar derivation can be carried out in case when I_B is also a variable. We will see an example in the temperature aware design in the next section.

In the following sections, we exploit this adaptive circuit design technique for two example applications, one for reducing temperature sensitivity and the other for mitigating the impact of circuit aging. Both of these examples demonstrate how the proposed adaptive circuit design can help the system achieve its optimum operating point and maximize the system computation efficiency. While temperature- and reliability-aware designs are in themselves important topics for reliable sustainable computing, we mainly choose these two examples to show the ability and versatility of the proposed scheme to adapt to different undesirable environmental factors and mechanisms, thus enabling typical case-based design. One should always remember that *green computing is not necessarily equivalent to low power computing*. The amount of consumed *energy* (or the produced carbon footprint for a given computation load) is the ultimate metric for measuring the 'green' part. Therefore, maintaining a circuit or system at its optimal operating point is the key to exploit every bit of the consumed energy for useful computation.

THRESHOLD COMPENSATION FOR TEMPERATURE VARIABILITY

Now we demonstrate how the proposed circuit technique can be used for maintaining an optimal I_{ON} to I_{OFF} ratio for MOS transistors in the presence of temperature variation. While energy, power, temperature, and temperature variations are all important issues for a system built from semiconductor chips, which in turn are built from transistors, it is important to distinguish among these concepts since the design challenges and issues are not always the same, thus cannot always be tackled together. As we will show in the next section, distinguishing these concepts is critical in identifying and directly targeting the challenge for reducing carbon footprint of computation and bringing forth green computing.

Temperature Variation and Why It Is Important

Thanks to the increase in circuit density and clock speed, power and thermal issues have been upgraded to first tier concerns in modern VLSI design, from chips in high performance servers to low power handheld devices (Chandrakasan, Sheng, & Brodersen, 1992). The first thought that usually comes to mind when thinking about thermal issues is probably power consumption, which is indeed a critical concern especially for battery supported, mobile and handheld consumer electronics and devices. The most popular examples are smart phones, tablets, laptops, portable media players, GPS (global positioning system) devices, etc. This concern about power motivated the whole low power design research initiative in both academia and industry, for good reasons.

However, more precisely speaking, for the above applications what we really need should be called ‘low energy’ design. Recall that energy is the product of power and time. For the given amount of computation (or device usage), simply lowering power certainly helps in reducing thermal hazards, but does not necessarily reduce energy consumption (or the battery usage). In reality, most low power designs or temperature-aware techniques involve lowering circuit speed and performance, thus increase the operation time for a given user task (Lu, Lach, Stan, & Skadron, 2005). For example, the most common and conceptually simple low power technique that has been long known in academia but has only been slowly adopted by industry in recent years, DVFS (Dynamic Voltage and Frequency Scaling), directly plays with this tradeoff between power and time by dynamically increasing or decreasing voltage and frequency at runtime for different applications. Since lower voltage and frequency immediately reduces power, but also decreases circuit speed (equivalently increasing computation time), therefore applying low power techniques like DVFS blindly does not necessarily reduce the energy consumption and increase battery time, but may actually exhaust their battery faster due the extra area and power overhead from the added DVFS management circuitry, in addition to possibly resulting in an unfriendly user experience due to the slower response time. As another example, parallelism can bring back the performance loss due to voltage scaling in low power design (Chandrakasan, Sheng, & Brodersen, 1992), but with leakage power rising rapidly in modern semiconductor technologies, the total energy consumption may still increase due to the extra transistors and area. Therefore, designers should always double check if they are using the right metric for their purpose, and if the metric is indeed improved after everything is considered. In the circuit design community, the Energy-Delay Product (EDP) is often used to evaluate the optimality, since it takes into consideration both energy consumption and system performance.

In this section, we focus on a less obvious, yet equivalently important, challenge hidden in the ‘power wall’ (a terminology from Asanovic, et al., 2006) temperature. Temperature is closely related to circuit performance and power. On one hand, the temperature is among the most important environmental factors, which can impact both the performance and power consumption (including dynamic runtime power and static leakage power) of integrated circuits. On the other hand, the temperature can also be affected by the power consumption of the circuit itself since all the power consumed by the circuit ultimately transforms into heat, according to the low of conservation of energy. In fact, since an elevated temperature not only hurts circuit performance but may also lead to a phenomenon called 'thermal runaway' (caused by a positive feedback of temperature and leakage power), which permanently destroy a semiconductor microchip, keeping circuit operational temperature in the specified range has become a critical concern, from integrated circuit packaging to IT facility maintenance, and is very cost-sensitive. At the macroscale level, the cooling facility is a significant portion of IT expense. The old cooling method through ventilation is no longer effective. Nowadays, high performance servers or compute farms need to use sophisticated air-conditioning schemes. At the microscale, the on-chip cooling systems like heat sink and spreader designs are getting fancier and more complicated, at the expense of cost that goes to every single chip, and has become a significant portion of the cost of chip manufacturing. Therefore, it is very important to distinguish between the two related concepts, i.e., low power design and temperature aware design.

While keeping the circuit out of dangerous temperature zones is vital for its health, keeping the circuit at optimal temperature can maximize its efficiency, since the variation of temperature can change the optimal operating point of a circuit. Unfortunately, while it may be possible to maintain the ambient or average temperature, it is almost impossible to maintain the temperature for every transistor on the circuit at a certain value, because of on-chip temperature variations. On-chip temperature variations can happen both spatially across different function blocks of the die, and temporally over time when the system runs different applications. Spatial temperature variations occur when some parts of the circuit consume more power, leading to 'hot spots' on the die while other parts are less active and their temperature is relatively lower. Temporal temperature variations occur when the function of the units on the die are switched on or off, changing the power consumption of the processor, thus changing the temperature of the chip over the time. The sources of temperature variations include ambient temperature and heat generated by circuit itself. Temperature variations are usually defined using 'temperature maps' (temporal and spatial distribution of temperature) which are either directly measured or indirectly derived from 'power maps' (temporal and spatial distribution of power). The calculation of temperature maps from power maps has been investigated by many research groups and various thermal simulators have been proposed for this purpose. On the other hand, temperature variation is also a good indicator of power distribution (Qi, Meyer, Huang, Ribando, Skadron, & Stan, 2010). In practice, circuits on a microprocessor die can experience a wide temperature range of 0-70°C for commercial applications and –55-125°C for military applications.

Due to the importance of temperature and the difficulty to control the temperature for every transistor because of on-chip temperature variations, researchers have proposed different techniques to compensate for temperature variation to reduce its impact (Park, et al., 1995). Most of them focus on scaling and exploiting the "design knobs" of V_{TH} and $V_{DD.}$ For example, using fixed-voltage and scaled-V_{TH} techniques, two sub-1V MOS-based reference current generators were proposed in (Tang, Narendra, & De, 2003) to provide immunity against both temperature and process variations. A dynamic voltage scaling technique based on the temperature variation was proposed in Shakeri and Meindl (2003) to reduce power consumption. In the following, we present the design of a temperature-aware circuit using the proposed threshold compensation technique to dynamically track temperature variations and reduce its impact on circuit performance.

Dependence of I_{ON} and I_{OFF} Current on Temperature

To explain the effect of temperature variation on circuit performance, we first show the dependence of the I_{ON} and I_{OFF} current on temperature. Here we are mainly concerned with digital circuits, which form the dominant portion of modern VLSI system and consume most energy and chip area. The input voltage is either close to V_{DD} (called '1' or ON state) or close to ground (called '0' or OFF state). When in the 'ON' state, a transistor operates in the saturation region. When in the 'OFF' state, a transistor operates in the subthreshold region).

In the ON state, the drain current of a transistor I_{ON} can be expressed by the α power law model (Sakurai & Newton, 1991) as shown in Equation (1). For simplicity we rewrite Equation (1) as:

$$I_{ON} \propto \mu(T)(V_{DD} - V_{TH}(T)) \quad (10)$$

where α is velocity saturation index whose typical value is 2 for long channel MOSFETs and approaches 1 for short channel MOSFET. For current technology α is around 1.3. Notice that the threshold voltage $V_{TH}(T)$ and the mobility $\mu(T)$ themselves also have temperature dependence that can be expressed as follows (Bellaouar, Fridi, Elmasry, & Itoh, 1998):

$$V_{TH}(T) = V_{TH}(T_0) - K(T - T_0) \quad (11)$$

$$\mu(T) = \mu(T_0)\left(\frac{T}{T_0}\right)^{-m} \quad (12)$$

Here T_0 is room temperature (300 Kelvin); K is a threshold voltage temperature coefficient whose typical value is 2.5mV/K and m is mobility temperature exponent whose typical value is 1.5.

According to (11) and (12), it can be seen that both mobility μ and threshold voltage V_{TH} has negative dependence on temperature. Moreover, from (10) we can see that μ and V_{TH} have the opposite impacts on the I_{ON}, thus it is not immediately obvious how I_{ON} changes across a range of temperature values. Figure 2 shows the temperature dependence of I_{ON} for a NMOS transistor with different supply voltage. Based on the analysis in (Park, et al., 1995), if the supply voltage $V_{DD} >> V_{TH}$, I_{ON} is dominantly controlled by the carrier mobility μ and the I_{ON} decreases as temperature increases; However, as V_{DD} approaches V_{TH}, the effect of the V_{TH} shift with temperature becomes stronger and I_{ON} increases with temperature. At a certain V_{DD} value (V_{DD} =0.7 in our chosen technology node as shown in Figure 2), I_{ON} does not vary with temperature. This voltage is called Zero-Temperature Coefficient (ZTC) supply voltage. In current technologies, V_{DD} is still much larger than V_{TH}, therefore temperature has a negative impact on I_{ON} current, and circuit performance deteriorates at higher temperatures.

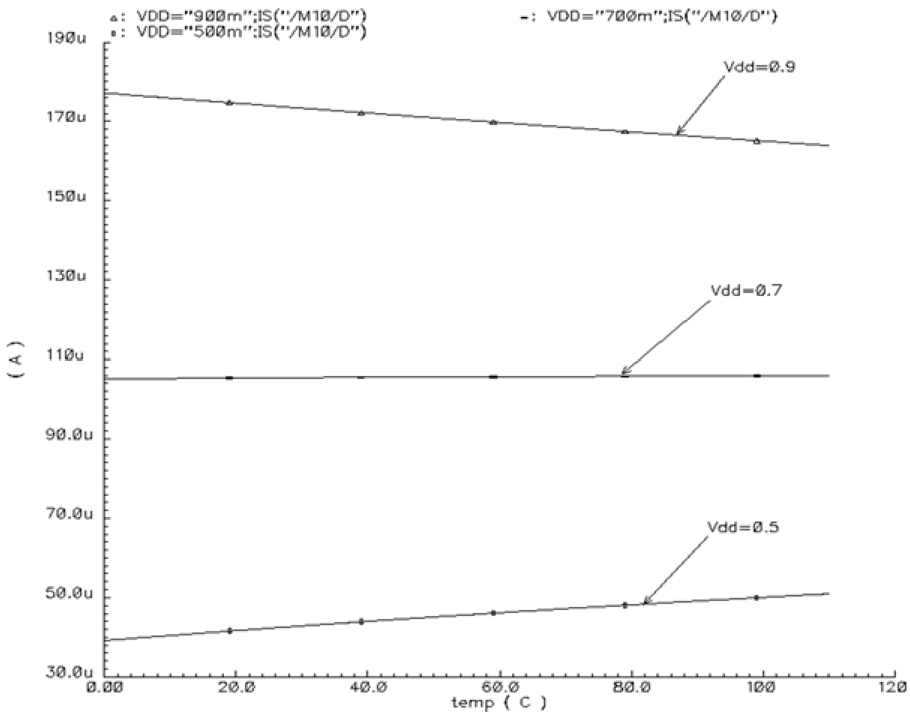
When a transistor is turned off, the drain current I_{OFF} is essentially the subthreshold current or leakage current, which can be expressed as:

$$I_{OFF} \propto \mu(T)\exp\frac{V_{GS} - V_{TH}(T)}{nV_t}\left(1 - \exp\left(-\frac{V_{DS}}{V_t}\right)\right) \quad (13)$$

Here n called the subthreshold slope coefficient and V_t=kT/q is the thermal voltage (which is about 26mV at room temperature). Equation (13) shows that I_{OFF} has an exponential dependence on V_t, so I_{OFF} increases exponentially as temperature increases.

Therefore, the effect of temperature variations on I_{OFF} is more significant than its effect on I_{ON}. This leads to a wide range of I_{ON} to I_{OFF} ratio values across the specified temperature range. The I_{ON} to I_{OFF} ratio characterizes the difference between the 'ON' state and the 'OFF' state, thus this value directly affects the robustness of a circuit (for example, a large I_{ON} to I_{OFF} ratio leads to a large noise margin, and a small I_{ON} to I_{OFF} ratio means more susceptibility to power noises and soft errors). This difference

Figure 2. Temperature dependence of I_{ON} with different supply voltages (Zhang & Stan, 2007 ©2007, ACM, Inc. Used with permission)



must be kept greater than a certain minimum, and the system must be designed to operate at the worst case I_{ON} to I_{OFF} ratio (which in turn implicates the upper bound of the operating temperature range). In a worst case-based design, the circuit will have to be over-designed to accommodate for the worst case. The over-designed circuit will not only incur power and area overhead, but also operate at non-optimal points whenever and wherever the temperature is lower than the upper-bound temperature. For this reason, we adapt the aforementioned threshold compensation scheme to dynamically track the on chip temperature variations and reduce the I_{ON} to I_{OFF} ratio variations across a wide temperature range.

Temperature-Aware Adaptive-Biasing Technique

In order to keep the ratio of I_{ON} to I_{OFF} constant across the temperature range, we can apply the threshold compensation scheme. Recall from Equation (5), positive bias voltage (or forward body bias) decreases the V_{TH} while negative bias voltage (or reverse body bias) increases V_{TH}. The key of the design is to adaptively bias the body voltage so that it tracks the circuit temperature thus produce a relatively constant I_{ON} to I_{OFF} ratio. There are two points we need to consider in the circuit implementation of the general compensation scheme.

First, as we have seen from the analysis in the last section, the temperature variations affect both I_{ON} and I_{OFF}, and their dependence on temperature is quite different. Therefore, here it is not enough to simply design the I_A from upper part of the circuit to follow a constant current I_B in the lower part, as shown in Figure 1. Instead, both I_A and I_B are changing with temperature and at different rates, and we need to design for the I_A to I_B ratio (or effectively, I_{ON} to I_{OFF} ratio). Therefore, we need to carefully design the parameters for both the upper and lower portions of the circuit to obtain a dynamic balance of the two, and produce appropriate output body bias voltages to achieve the temperature compensation for the I_{ON} to I_{OFF} ratio.

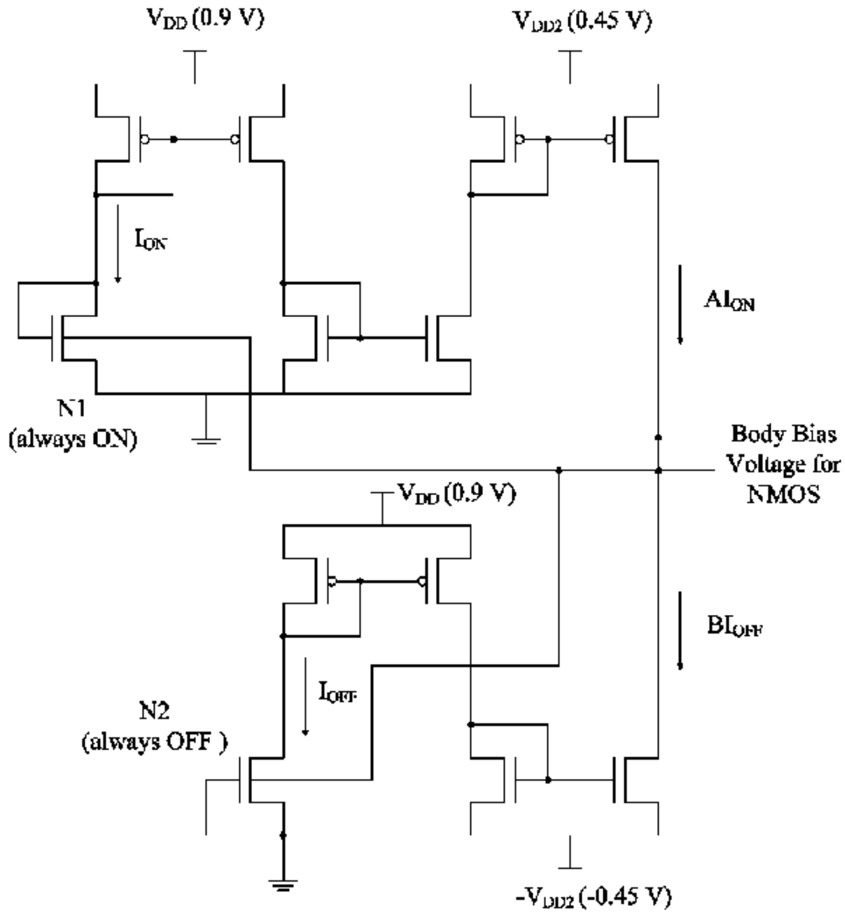
Second, the most common CMOS (Complementary MOS) based digital circuits have a pull up network comprised of solely PMOS transistors and a pull down network comprised of solely NMOS transistors. The strength of the pull up network affects the rising edge of the output signal, while the pull down network affects the falling edge of the output signal. To reduce the temperature variation induced slew rate changes, both pull up and pull down network need adaptive body biasing to reduce their current change due to the temperature variations. Since NMOS and PMOS transistors have different body bias voltages (the body of NMOS transistor is usually close to '0' or ground, while the body of PMOS transistor is usually close to '1' or power supply voltage), we need to design two sets of body bias circuits to supply different body bias voltages to both transistor types to adjust their drain currents separately.

Figure 3 shows the adaptive body biasing circuit implementation for NMOS transistor temperature compensation. To understand the circuit, we can first view it in two parts, i.e., the upper and lower branches. These two parts monitor *ION* and *IOFF* respectively. In the upper part, N_1 is ON (since its gate is connected to drain) and thus its current is a proxy for I_{ON}. In the lower part, NMOS transistor N_2 is OFF (since its gate is connected to the source) and thus its current is a proxy for I_{OFF}. Then for each part, current mirrors are sized with fixed gains *A* and *B* to separately amplify the *ION* and *IOFF* currents. At the final output the circuit is in equilibrium when $A \times I_{ON} = B \times I_{OFF}$ (again, *A* and *B* are the gains for the ON and OFF branches). The circuit is designed such that when the temperature equals a certain value, the output voltage in the equilibrium state is close to zero, thus the circuit does not provide any bias voltage to the NMOS transistors in the system at this temperature. Then, when the temperature changes due to various reasons including ambient environmental change or on chip power dissipation, the circuit will output a non-zero bias voltage to balance the *ION* and *IOFF* currents and reduce the temperature sensitivity of I_{ON} to I_{OFF} ratio. In this work this zero biasing temperature is set at the middle point of the specified temperature range, but it could also be set to other values, e.g. to the expected working temperature.

To see how the temperature sensitivity can be reduced, first consider the case when the temperature becomes lower than the zero biasing temperature. Since I_{OFF} decreases much faster than I_{ON}, this leads to an imbalance of $A \times I_{ON} > B \times I_{OFF}$. This imbalance will move the output voltage to be more positive, thus providing a positive bias or forward body bias voltage for NMOS. The forward bias in turn reduces *VTH* and increases I_{OFF} to bring the circuit back to equilibrium. On the other hand, if the temperature is larger than the zero biasing temperature, the faster increase in I_{OFF} leads to an imbalance $A \times ION < B \times I_{OFF}$, which results in a negative output (or reverse body bias for NMOS) which increases *VTH* and decreases I_{OFF} to bring the circuit back to equilibrium.

We note that the threshold compensation circuit requires two additional supply voltages (+0.45V and -0.45V in the implementation shown in Figure 3, but may vary depending on technology and design parameters) to generate the output bias voltage, which would be implemented either through external power pads or internal charge pumps. Moreoever, the value of the output voltage can be used as an indication of the temperature. In other words, each value of V_{OUT} corresponds to a certain temperature value, so with a look up table this circuit can also acts as an on-chip temperature sensor. Finally, implementing the threshold compensation for the PMOS transistor to reduce the temperature sensitivity of its I_{ON} to I_{OFF} ratio is very similar (mirror) to the NMOS one.

Figure 3. Threshold compensation circuit for NMOS to reduce temperature sensitvity of I_{ON} to I_{OFF} ratio (Adatepd from Zhang & Stan, 2007)

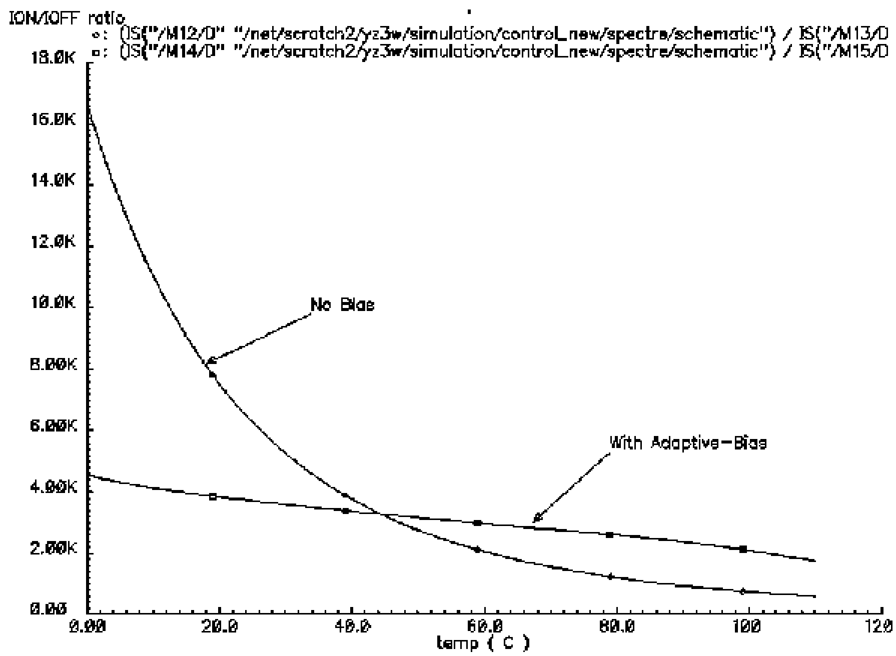


Temperature Sensitivity Reduction of I_{ON} to I_{OFF} Ratio

The threshold compensation circuit shown in Figure 3 has been designed in the openly downloadable 70nm Predictive Technology (PTM) (Zhao & Cao, 2006) with the nominal supply voltage of 0.9V. All the simulations in the following sections are also performed with this technology. The output biasing voltage varies from 320mV to –280mV across the temperature range (0 to 110°C). The zero point is set in the middle of the operation temperature range (55°C). We also run Monte Carlo simulations to see the impact of parameter variations. The result shows ~10% standard deviations at high bound and low bound range of output bias voltage.

Figure 4 shows the I_{ON} to I_{OFF} ratios for a single NMOS with and without threshold compensation. The ratio varies from 16600 to 500 without biasing when the temperature varies from 0°C to 110°C. As can be seen, at low temperatures *IOFF* is reduced significantly and circuits operate at a larger *ION/IOFF* ratio than necessary. This leads to sub-optimal performance. By applying forward body biasing to reduce the threshold voltage, the *ION/IOFF* ratio can be brought back to the nominal value. As can be seen from Figure 4, with the proposed compensation scheme the ratio is kept within a much narrower range and only

Figure 4. ION/IOFF ratio for a NMOS transistor with and without threshold compensation



varies from 4500 to 1800. Also, due to the smaller threshold voltage at low temperature, the circuits operate at higher speed. Therefore, in this case the performance can actually be improved with a scheme like Dynamic Voltage and Frequency Scaling (DVFS), which takes advantage of higher circuit speed to reduce computation time with equal or less energy consumption for a given task.

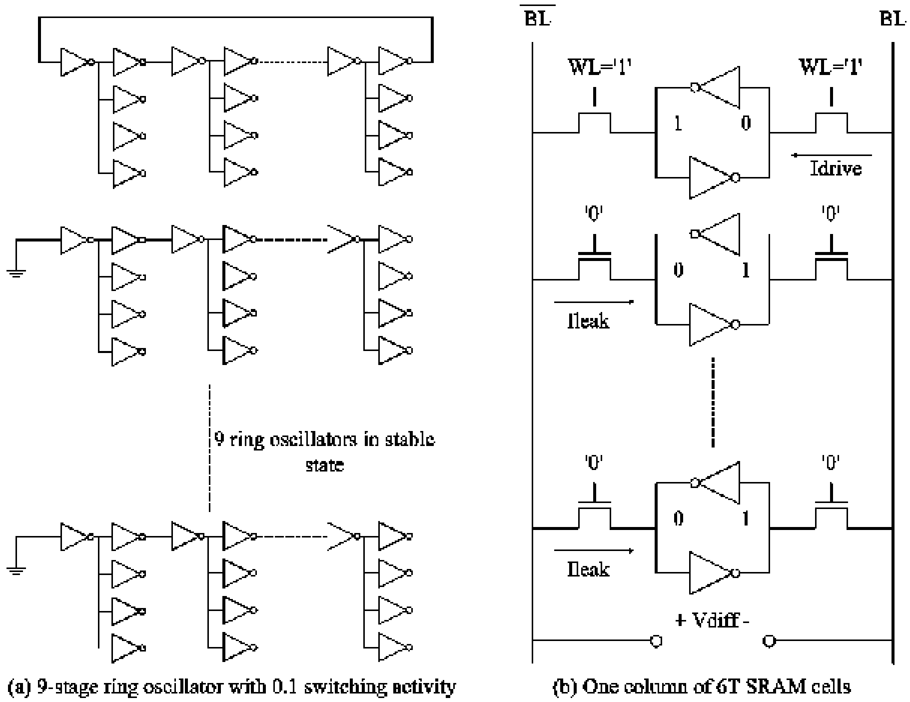
Temperature Compensation for Combinational Circuitry

We first apply the compensation scheme for logic circuitry. A circuit that represents a generic logic circuit is shown in Figure 5(a). The first block in the test circuit is a nine-stage, Fan-Out of 4 (FO4) ring oscillator. Then other nine identical inverter chains are constructed and put in a static state. Together they form a system with an equivalent switching activity factor of 0.1. Since at different supply voltages I_{ON} has a different temperature dependence, two sets of simulations are run at V_{DD}=0.9V (negative dependence) and V_{DD}=0.5V (positive dependence).

The simulation results are shown in Figure 6. There are two other baseline settings in our simulation. "NO BIAS 1" represents test circuit without threshold compensation. "NO BIAS 2" represents a test circuit without threshold compensation but still achieving the same worst-case I_{ON}/I_{OFF} ratio at 110°C (in order to do this we have to increase default threshold voltage to match the actual threshold voltage with threshold compensation at 110°C). Finally, "TAC-ABB" represents the temperature aware circuit with adaptive body biasing circuitry for threshold compensation.

Figure 6a shows the comparison of I_{ON}/I_{OFF} ratio between the three settings for different supply voltages. Here I_{ON} current is the average drain current in the ON state and I_{OFF} the average leakage current in the OFF state. With threshold compensation, the ratio has much less temperature sensitivity and remains almost constant across the temperature range for both supply voltages. Furthermore, "NO BIAS 1" has lower worst-case $I_{ON}/$

Figure 5. Two tests circuit for temperature variability compensation: (a) represents a combinational circuit and (b) represents a memory circuit (figures from Zhang & Stan, 2007)

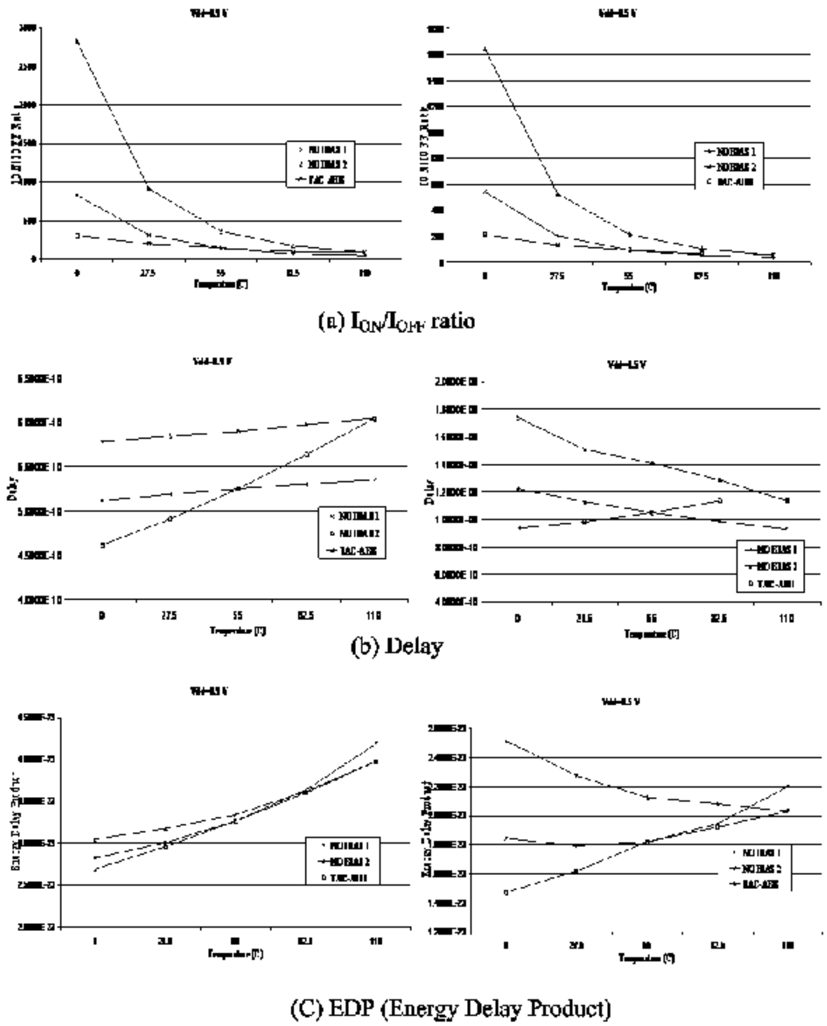


I_{OFF} ratio compared to other two settings. "NO BIAS 2" has the same worst-case I_{ON}/I_{OFF} ratio as the "TAC-ABB" but the ratio at lower temperatures is significantly higher than the other two settings.

Figure 6b shows the comparison of delay between three settings for different supply voltages. At low temperatures, the system with threshold compensation operates at a higher frequency and lower delay thus offering a higher performance compared with both "NO BIAS 1" and "NO BIAS 2." At high temperatures, the worst case-based "NO BIAS 2" design offers the same I_{ON}/I_{OFF} ratio. However, this comes at the expense of longer delay at other temperatures since it has higher V_{TH} value than the other designs. For non-temperature-aware designs like "NO BIAS 1," the leakage current increases drastically, which can cause reduced noise margins, potential thermal hazards and increased leakage power. With the proposed threshold compensation scheme, the system operates at a lower frequency at high temperatures to retain the system's reliability and stability. By operating at lower speed, the circuit also generates less power, which can help cool down the system. Moreover, with reverse bias, the NMOS transistors have higher V_{TH}, which further cuts down the leakage power. Therefore, the threshold compensation scheme can indeed help reliable and green computing.

On the other hand, the right hand side of Figure 6b shows that for lower values (here V_{DD}=0.5 V), the circuit speed could have a positive temperature dependence, i.e., the circuit becomes faster at higher temperature. This reverse temperature dependence usually happens for low power supply voltages (Park, et al., 1995). This phenomenon can lead to a positive feedback, in which the circuit runs faster at higher temperatures, and generates more power to further increase the temperature. This positive feedback loop may cause serious reliability problem. Again, the dynamic threshold

Figure 6. Comparing I_{ON}/I_{OFF} ratio, delay and energy delay product in three scenarios across the temperature range 0 °C - 110 °C. Each coomparison is done for both (figure from Zhang & Stan, 2007).



compensation technique can throttle this mechanism to prevent potential thermal runaway.

Figure 6c shows the comparison of Energy Delay Product (EDP) between the three settings for different supply voltages. Overall EDP is improved with the threshold voltage for both high and low supply voltages. Compared to "NO BIAS 1," about 5% improvement is achieved at both 0°C and 110°C when V_{DD} = 0.9V; about 20% and 8% improvements are achieved at 0°C and 110°C respectively when V_{DD}=0.5. Compared with "NO BIAS 2" which by design has the same V_{TH} as the threshold compensation circuit, even larger improvements are achieved across all the temperature values. At 0°C EDP is reduced by 11.6% and 41.5% for V_{DD} =0.9V and 0.5V, respectively.

From the above simulation results, it can be observed that by reducing the temperature sensitivity of I_{ON}/I_{OFF} ratio, the threshold compensation scheme offers high performance and reliability. The scheme also achieves a smaller energy-delay product, which means that for the same task, it can reduce the energy without posing a penalty on performance. This is achieved by dynamically tracking the temperature and exploiting higher performance at low temperatures and reducing leakage power at high temperatures. When reverse temperature dependence happens, the same circuit can also throttle the system to prevent potential thermal runaway and unreliable conditions. In summary, the threshold compensation circuit helps the system remain in the optimal and most energy efficient condition.

Threshold Compensation for Memory Circuitry

Due to the large capacity and low switching activity, leakage is a much more critical metric in memory design than in combinational circuit design. However, leakage is important not only because it determines the static power consumption, but it can also affect SRAM stability. In this section, we show how the threshold compensation scheme also helps memory circuits. We implemented a 64-row × 32-column, standard 6T (six-transistor) SRAM array in the 70nm PTM model with 0.9 V supply voltage as an example. Figure 5b shows one column of the SRAM memory.

Assume the top cell which stores a "0" is being read. After both bit lines BL and $\overline{BL}$ are pre-charged to "1," the top access transistors are selected. BL is discharged while $\overline{BL}$ keeps its pre-charged value. After some time a certain amount of differential voltage (100mV in this example) is built up between the two bit lines and the sense amplifier for this column is activated to poll the stored value. Notice that the above description does not consider the influence of other cells in this column which all share the same bit lines. In the worst case, a "0" is stored in the top cell while all other cells are storing "1," therefore the leakage current from these cells results in a voltage drop on $\overline{BL}$. This voltage drop not only deteriorates the access time (thus memory performance) but also has an impact on the cell stability and noise margin.

At high temperatures, the leakage current *IOFF* increases faster than *ION*, thus by reverse body biasing and increasing *VTH*, *IOFF* can be reduced, thereby decreasing access time and increasing noise margin. At low temperatures, *IOFF* decreases quickly and the read current (*ION*) becomes the dominant factor of the read speed. Forward body biasing can reduce V_{TH}, increase *ION* and reduce the access time. Therefore threshold compensation can improve SRAM performance across whole temperature range. Notice that this is different from the combinational circuit case, where reverse body bias increases delay and throttles circuit speed.

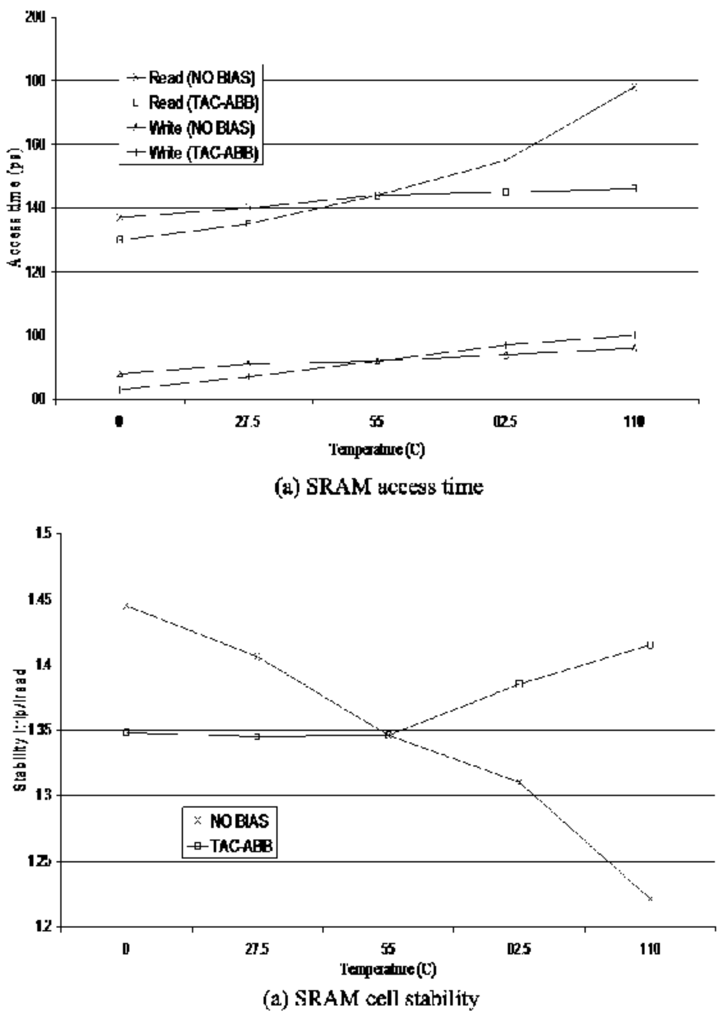
Figure 7a shows the performance comparison for SRAM. By applying threshold compensation, 5.1% and 18% read time improvement are achieved at low and high temperatures respectively. Write time increases slightly at high temperatures, but write is not a critical operation for memory access since it has a large noise margin, so this penalty is usually acceptable.

We use the method proposed in Hamzaoglu, et al. (2002) to measure the 6T SRAM cell stability. Figure 7(b) shows the stability comparison. At high temperature with threshold compensation, the SRAM cell stability is improved by 15.8% and worst-case cell stability is improved by 10%. This comes with a penalty of less stability at low temperatures. However, this penalty can be easily justified since most reliability problems happen in high temperatures. This increasing stability with rising of the temperature is especially suitable for high performance applications, which usually suffer from degraded performance and stability in elevated temperatures.

THRESHOLD COMPENSATION FOR NBTI

The previous examples have demonstrated how the threshold compensation circuits can be used to improve circuit performance, energy efficiency

Figure 7. Improvement of both performance and stability for SRAM cells by threshold compensation (figure from Zhang & Stan, 2007)



and thermal reliability by reducing temperature variability for both combinational and memory circuits. In this section, we show another area where the threshold compensation scheme can find applications.

The Circuit Also Ages

Even as it continues unabated, technology scaling hardly comes for free any more. Reliability has become a major concern and poses hefty penalties on future scaling (Borkar, 2005). Reliability is usually measured by MTTF (Mean Time to Failure), or failure rate (Hennessy & Patterson, 2003). However, different reliability hazards impact circuits in very different ways and need different ways of handling. For example, soft errors (also called single event upsets or SEU) occur almost purely randomly and cause random transient failures. Aging mechanisms like NBTI (Negative Bias Temperature Instability), HCI (Hot Carrier Injection), EM (Electro-Migration), TDDB (Time Dependent Dielectric Breakdown), and thermal cycling are more tractable, as they usually appear as a result of a certain amount of circuit functioning, and ultimately limit the system lifetime. Interestingly some aging effects are reversible. For instance, 'healing effects' exist for both NBTI (Bhardwaj, Wang, Vattikonda, Cao, & Vrudhula, 2006) and EM (Blaauw, Oh, Zolotov, & Dasgupta, 2003).

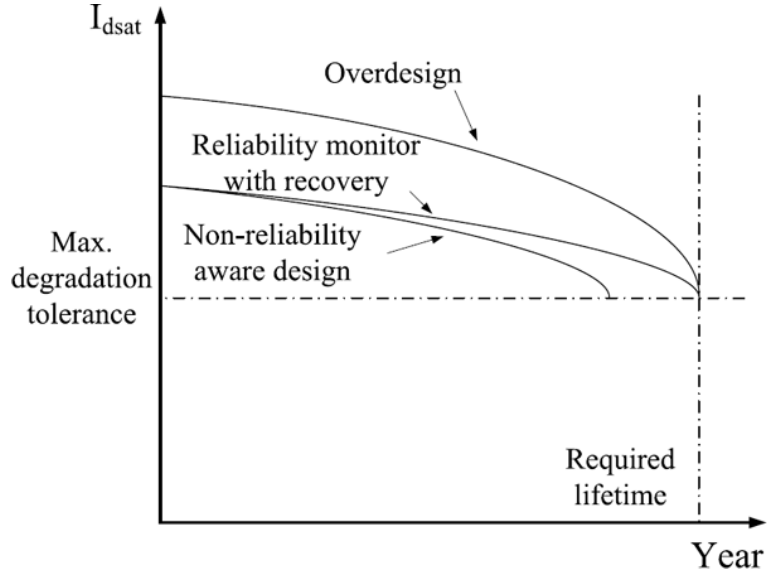
At the system level, reliability-aware design schemes, including Dynamic Reliability Management (DRM) and reliability banking, have been proposed (Karl, Blaauw, Sylvester, & Mudge, 2006; Lu, Lach, Stan, & Skadron, 2005). In a general sense, Dynamic Thermal Management (DTM) is also a reliability aware design method that targets temperature cycling (thermal shock). At the circuit level, redundancy based designs like TMR (Triple Modular Redundancy) and Razor (Ernst, et al., 2005) were introduced to prevent local circuit failures from propagating in the system. 'Circuit hardening' (Zhou & Mohanram, 2006), among other techniques that provide extra margin by overdesigning, is often adopted as well. Redundancy and overdesigning are general methods to deal with most reliability problems, but they could also lead to large area and power overhead by duplicating, triplicating, or upsizing transistors by a significant portion. Accurate reliability models and time consuming reliability simulations, which are not commonly found in current design kits and design tools, may also be needed. In addition, it can be difficult to calculate the required margin for desired lifetime if the realistic input signal pattern is unknown, since the switching activity directly decides the amount of NBTI stress for each PMOS.

It has been suggested that online monitoring with in-situ sensors might be more effective in dealing with reliability problems (Mishra, Pecht, & Goodman, 2002; Mitra, 2007). The idea is to track circuit degradation in real time and issue advance 'warnings' before the system starts to malfunction. A warning indicates either maintenance ('healing,' if possible) or replacement is required. Figure 8 compares the different impact of in-situ sensors and overdesigning on circuit lifetime. In contrast to the above worst-case design with excessive margins, reliability monitors enable typical-case designs. In this chapter, we propose a reliability monitor for NBTI that not only tracks the degradation online, but also *adaptively* heals the degradation and prevents it from harming normal system functioning. In advanced technologies with thin gate oxide, NBTI is predicted to dominate other reliability mechanisms like HCI in system lifetime (Bhardwaj, Wang, Vattikonda, Cao, & Vrudhula, 2006). Although the monitor here is designed specifically for NBTI, it could also be adapted for other reliability hazards as long as they can be monitored or mitigated through transistor threshold voltage, e.g. HCI and thermal hazards.

NBTI Modeling

NBTI becomes a major concern in modern CMOS technologies with thin gate oxide due to its exponential dependence on oxide thickness and temperature (Chen, et al., 2003). NBTI stress occurs in PMOS transistors with negative gate to source voltages, which is often the case in CMOS circuits where PMOS serves as the header device. The most notable aging effect from NBTI is the increase in absolute value of the threshold voltage V_{TH} for PMOS, which contributes to decreased transconductance and weaker driving capability. Although similar mechanisms exist for NMOS, i.e., PBTI (Positive Body Bias Temperature Instability), NBTI imposes a much shorter lifetime compared to PBTI and other major reliability concerns, and has becomes a very active research topic recently (Bhardwaj, Wang, Vattikonda, Cao, & Vrudhula, 2006; Kang, Kim, Islam, Alam, & Roy, 2007; Kang, Kufluoglu, Alam, & Roy, 2007; Paul, Kang, Kufluoglu, Alam, & Roy, 2006; Vattikonda, Wang, & Cao, 2006; Wang, et al., 2007). While this phenomenon has been known for more than 40 years in the device area, the underlying physics is still being studied under various circuit operating conditions (Massey, 2004), and can be

Figure 8. Circuit performance degradation comparison with different reliability aware designs (Qi & Stan, 2008 ©2008, ACM, Inc. Used with permission.)



more complicated when coupled with process variation (Agostinelli, et al., 2005).

Nonetheless, it is well understood that NBTI stress and recovery phases involve the diffusion and back-diffusion of H and H_2 in the oxide, the Poly-Si and Si. The following equations are from the newest NBTI model from PTM (Zhao & Cao, 2006), also explained in Bhardwaj, Wang, Vattikonda, Cao, and Vrudhula (2006):

$$\Delta V_{th} = \begin{cases} \sqrt{K_v^2 (t-t_0)^{0.5} + \Delta V_{th1}^2} + \delta_v & \text{(stress)} \\ (\Delta \mathrm{V}_{th2} - \delta_v)\left[1 - \sqrt{\eta (t-t_0)/t}\right] & \text{(recovery)} \end{cases} \tag{14}$$

where:

$$K_v = A\,T_{ox}\sqrt{C_{ox}(V_{gs}-V_{th})}\exp\left(\frac{E_{ox}}{E_0}\right) \left[1 - \frac{V_{ds}}{\alpha(V_{gs}-V_{th})}\right]\exp\left(-\frac{E_{ox}}{E_0}\right)$$

and:

$$E_{ox} = (V_{gs} - V_{th}) / T_{ox} \tag{15}$$

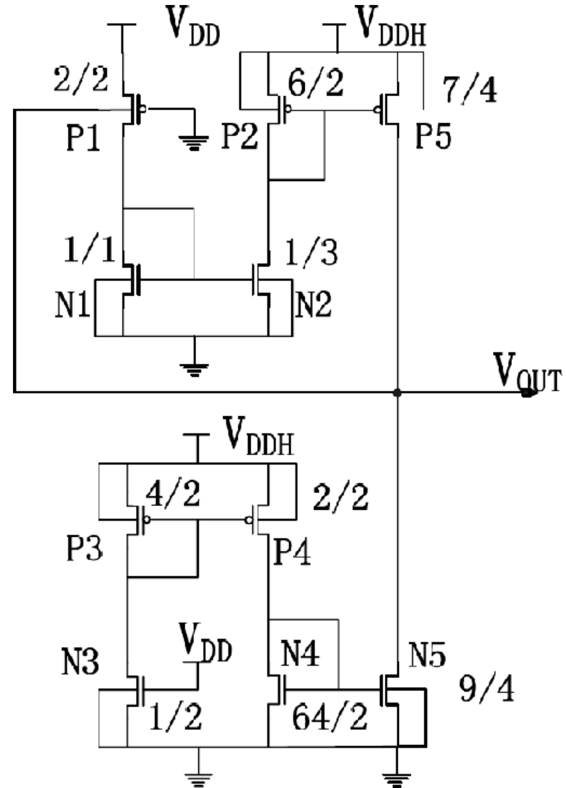
Default values for different PTM technology nodes can be found on the PTM website (PTM Technology). In the above model ΔV_{TH} is proportional to $t^{1/4}$, where t denotes time. Depending on whether H or H_2 dominates the diffusion, the dependence of ΔV_{TH} and t could vary from $t^{1/4}$ to $t^{1/6}$ in Bhardwaj, Wang, Vattikonda, Cao, and Vrudhula (2006), Kang, Kufluoglu, Alam, and Roy (2007), and Kumar, Kim, and Sapatnekar (2006).

Body Bias Circuit

Figure 9 shows the proposed threshold compensation scheme adapted for NBTI monitoring and mitigation. This circuit shares many similarities with the one used for reducing temperature variability shown in Figure 3. The circuit also can also be divided into an upper part and a lower part. In the upper part, transistor $P1$ is acting as a proxy for PMOS transistors experiencing NBTI stress. Transistors *N*1 and *N*2 serve as a current mirror which track the drain current of *P*1 and also provides some current gain. Similarly, *P*2 and *P*5 form the second current mirror that further tracks and amplifies the drain current. The lower part of the circuit also consists of current mirror circuits, except that now the current in the first stage is determined by the NMOS transistor *N*3, whose gate is constantly tied to V_{DD} and is always on. After two stages of current amplification through current mirror pairs *P*3/*P*4 and *N*4/*N*5, the amplified drain current of *N*3 eventually becomes the drain current of *N*5. Since the drain currents of *N*5 and *P*5 have to be the same, this balance will decide the output voltage V_{OUT}, which is then used to body-bias PMOS transistors of the actual circuit. The output is also fed back to *P*1 to form a closed loop amplifier. Similar as the circuit used for temperature variability, the magnitude of the biasing voltage can also be used as a metric to characterize the NBTI stress amount and the circuit can act as an on chip NBTI monitor.

However, there are some important differences between the design for NBTI in Figure 9 and the design for temperature variability in Figure 3. First, the gate of transistor *P*1 is tied to ground so V_{GS} = - V_{DD} for *P*1 all the time. V_{DD} is a voltage equal to the power supply voltage in the actual circuit to be monitored. Since each PMOS transistor will have its own switching activity, and most of them usually have some periods of stress relaxation when the input gate voltage turns high and shuts off the PMOS, this connection makes *P*1 subjected to the most severe NBTI stress compared to PMOS transistors in the actual circuit. The introduced pessimistic assumption can be dealt with at the circuit level, as will be discussed later. This is different from the case of temperature variability, where temperatures of transistors in the same area are usually quite close, and the proxy transistor could either see a higher or lower temperature than the actual circuit it is monitoring.

Figure 9. Threshold compensation circuit for monitoring and mitigating NBTI. Sizing is shown for PTM 65nm technology with V_{DD} =1V. Feature size (65nm) is taken as unit length (Qi & Stan, 2008 ©2008, ACM, Inc. Used with permission.).



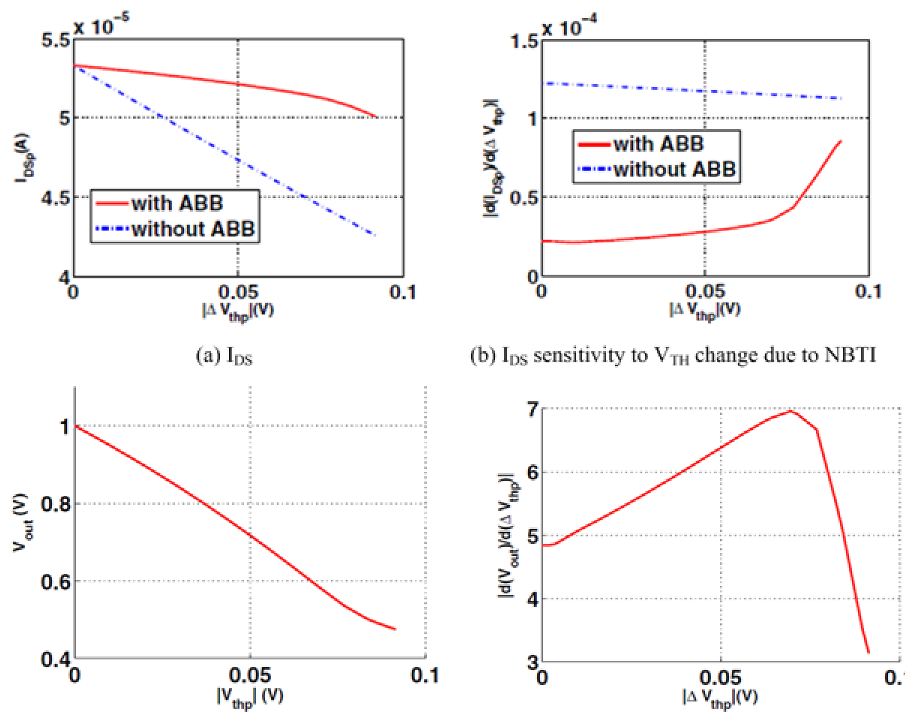
Second, there is only one power supply V_{DDH} in the last stage. In the temperature variability case, both positive and negative voltage supplies are needed since the starting point is at the middle and temperature can both rise and fall. On the other hand NBTI degradation only moves PMOS threshold voltage more negative. Therefore, only forward body biasing will be applied to mitigate this change. Since V_{OUT} should equal V_{DD} when no NBTI stress is applied, V_{DDH} needs to be higher than V_{DD} to provide full swing to the output body bias voltage. While this may pose extra design challenges and potential reliability problems, multi- V_{DD} design is not new in today's IC design flows and is well supported by CAD tools. Furthermore, body biasing (both forward and backward) has been adopted as a standard method to systematically reduce the impact of process variation (Tschanz, et al., 2002), which also implies that voltages higher than V_{DD} can be made available.

Finally, it is worth mentioning that at least on the first order ΔV_{TH} is independent of transistor sizing, as can be seen from Equation (14). On the other hand, the drain voltage will affect NBTI degradation. For PMOS stacks in the pull-up network in CMOS gates, the degradation of each PMOS gate can be different. However, the V_{DS} impact on ΔV_{TH} is rather insignificant, and stacks of PMOS gates are not recommended anyhow due to the slow speed.

Experimental Results

Just as in the temperature-aware design, we start from a simple test case. We first examine the drive current I_{DS} of two PMOS transistors to see how device degradation due to NBTI can be mitigated with the proposed scheme. The following simulation results are obtained with PTM 65nm models, $1V\ V_{DD}$ and $2V\ V_{DDH}$. The NBTI monitor circuit is sized as shown in Figure 9. Figure 10(a) shows how the I_{DS} of each transistor decreases as $|V_{TH}|$ increases. Here we assume that NBTI imposes exactly the same V_{TH} degradation on all PMOS. The reduced I_{DS} amounts to a larger time constant, weaker driving capability and longer charging time, which can eat-up timing margins and eventually lead to timing violations. Although typically NBTI lifetime can be mapped to ΔV_{TH}, what we would like to find out is how lifetime is extended by the reliability monitor. Therefore, we take a 5% I_{DS} reduction as the cut-off point for lifetime estimation and the ΔV_{TH} shown in all the figures does not take body biasing into account in order to isolate ΔV_{TH} caused by NBTI. This is

Figure 10. Current and output voltage change with threshold voltage drift, as well as their sensitivity to threshold voltage drift (Qi & Stan, 2008 ©2008, ACM, Inc. Used with permission.).



(a) I_{DS} (b) I_{DS} sensitivity to V_{TH} change due to NBTI

(c) V_{OUT} (d) V_{TH} sensitivity to V_{TH} change due to NBTI achieved by modifying parameters in the PTM model file during simulation.

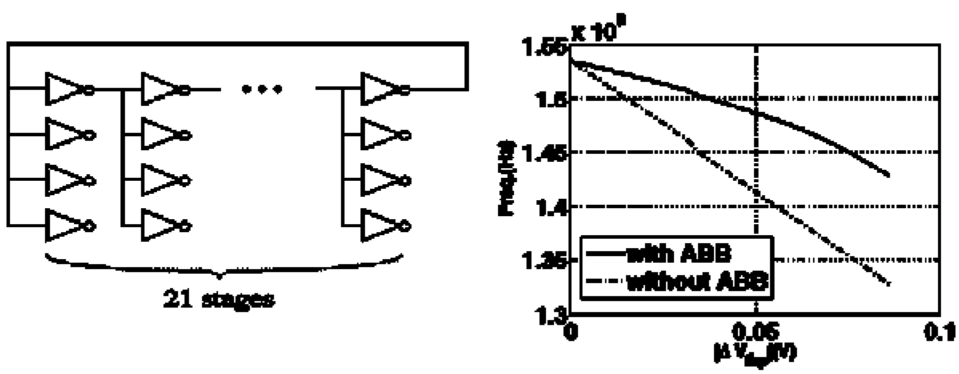
From Figure 10b, it can seen that with ABB, I_{DS} reduction in $P2$ is much slower than $P1$, which means P2 tolerates more NBTI stress and has longer lifetime. Indeed, the 5% ΔI_{DS} for $P2$ is reached when V_{TH} degrades almost 100mV from the unstressed value. In contrast I_{DS} in $P1$ drops 5% as soon as V_{TH} changes by only 25mV. Figure 10b compares the sensitivities of I_{DS} in both transistors. It can be seen that body biasing essentially reduces the sensitivity, especially in the range where the maximum reliability tolerance is not violated and the device is supposed to meet the performance requirement set by timing analysis.

Referring to the NBTI model from earlier sections, ΔV_{TH} depends on $t^{1/n}$ where n varies from 1/4 to 1/6. With the reliability monitor, the tolerance of ΔV_{TH} increases four times from 25mv to 100mv, which in turn is equivalent to a prolonged device lifetime 256 to 1296 times the original depending the effective n value in the NBTI model. However, notice that the above values are *not* representative of the actual circuit lifetime due to the assumed constant static stress. The above analysis is for static stress with no recovery phase during the whole lifetime. In practice very rarely will a PMOS gate be tied constantly to ground, which is the worst-case scenario. Moreover, device lifetime is also limited by temperature and other reliability hazards. Nonetheless, the improvement in lifetime demonstrates the effectiveness of the proposed NBTI monitor.

Figure 10c and 10d show the output voltage and the small signal gain of V_{OUT} to V_{TH}. The circuit is sized such that this gain is maximized within the lifetime requirement. This is an important guideline for this design. Another constraint is that the PMOS body bias voltage must equal V_{DD} at the beginning when any NBTI stress is applied, as it would be if the monitor did not exist. For different technologies, the circuit in Figure 9 needs to be resized accordingly to these requirements.

To show the effectiveness of the NBTI compensation for general circuits, we again resort to the ring oscillator structure. This time we use a twenty-one stage FO4 ring oscillator designed in PTM 65nm technology. Again, with the assumption that all PMOS transistors have the same NBTI stress and thus the same ΔV_{TH}, we monitor the frequency degradation. As shown in Figure 11, with the NBTI monitor, the circuit can tolerate almost twice ΔV_{TH}. However, in this case the lifetime cannot be simply translated from $t^{1/4}$ or $t^{1/6}$, since this is dynamic NBTI stress with relaxation and 'healing' within every clock period.

Figure 11. Threshold compensation slows down the frequency degradation of a 21 stage oscillator as PMOS threshold voltage being degraded by NBTI (Qi & Stan, 2008 ©2008, ACM, Inc. Used with permission.).



Actual energy savings from employing the proposed scheme relate to a number of factors, including the system specification, usage model, circuit-switching activity, operating environment, deployment of the proposed circuit, and the alternative approach when the proposed scheme is not adopted, etc. To make our analysis more straightforward, we still resort to the simple but representative ring oscillator in Figure 11. With a specification that states that the frequency cannot fall below 900MHz, this results in about 5% guardband, which translates to 0.08V threshold degradation. Without the proposed scheme, at 0.08V threshold degradation the circuit would be operating at 830MHz, thus violating the specification. Ignoring all second order factors, the maximum frequency is inversely proportional to the lesser of the charging and discharging cycles of the circuit. Since NBTI only affects PMOS, the charging cycle becomes the limiting factor. The charging cycle is proportional to the maximum of the current that a PMOS can supply. Therefore:

$$f \propto \frac{1}{t_{charging}} \propto I_{PMOS} \propto W_{PMOS} \qquad (16)$$

It is immediately clear that in order to maintain 900MHz when NBTI degrades the V_{TH} by 0.08V, one needs to add an up-front guarband of (900-830)/830 = 8.4% on the PMOS width, which translates to 8.4% extra energy consumption and area overhead.

Notice that the above is an over-simplistic first order estimation. First, overdesign doesn't need to be applied on every path. However, one cannot overdesign only the critical paths either. Identifying which paths to overdesign is a big challenge and requires extensive low-level circuit simulation based on accurate models from device-level to system-level (Saluja, Vijayakumar, Sootkaneung, & Yang, 2008). Second, the actual energy savings will be offset a little by the energy consumption of the threshold compensation circuit itself. The overhead incurred by the proposed circuit is largely affected by its deployment, which is discussed in the next section.

It should also be emphasized that the benefits of the proposed scheme go far beyond the immediate energy saving from the above guardband reduction. In fact, the more implicit energy savings can actually be more significant. By tolerating more aging degradation, the circuit improves the efficiency of analog circuits which often rely on precise matching and voltage biasing. It also improves noise characteristics and soft error immunity in memory systems. In contrast, without adaptive threshold compensation and merely adding guardband disrupts the normal design. For example, upsized PMOS transistors form a stronger pull-up network than the pull-down network, thus reducing noise margins. With the adaptive compensation scheme, the reduction of excessive noise and probability of failure lead to potential energy savings which would otherwise be incurred from engaging error correction systems or replacing failure part. For example, in a recent paper (Mostafa, et al., 2011) which proposes an alternative body bias scheme, our circuit is shown to improve write margins by 4.6%, and reduces the read and write failure probablity by 0.28% and 0.14% respectively for standard 6T SRAM banks. In fact the proposed circuit is also shown to be 1.2X smaller in area, and leads to 5.5% less leakage power than the ABB circuit proposed in Mostafa, et al. (2011).

DEPLOYMENT OF PROPOSED CIRCUIT

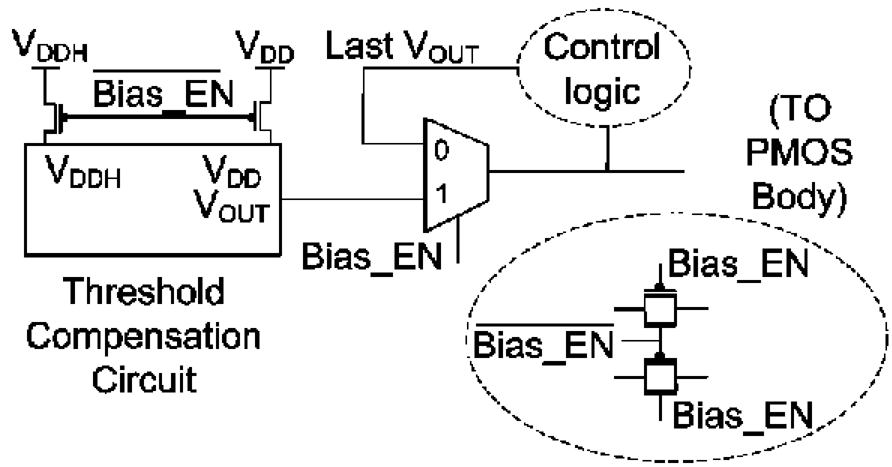
As shown in both applications, the proposed threshold compensation circuit serves dual purposes. It can act as an on chip sensor to monitor temperature or reliability effects, and at the same time reduce their impact on the system by compensating the threshold voltage change via adaptive body biasing. A common issue with on chip sensors is how to let them see the same environmental variables (temperature, NBTI, etc.) of the actual devices that they are monitoring. Each individual device can experience different temperature or stress, and it is impossible to assign one sensor to each device for the whole circuit. Process variations further complicate the problem. A more realistic and practical way is to allocate a sensor to a group of devices in their neighborhood in order to reduce the overhead and the impact of local process variation. The granularity at which the sensors are deployed reflects a tradeoff between overhead and accuracy, and the optimization can be assisted with a-priori knowledge of temperature or failure distribution.

As mentioned above, compensating for aging effects is more challenging than for temperature variability, since the former is based on switching activity which differs from device to device and ideally the compensation should have much finer granularity than for temperature variability. In the following we use the NBTI application as an example and provide some analysis and insight. In Figure 9 $P1$ is under constant NBTI stress, thus only representing the worst stress case. This pessimism may result in a situation where a PMOS is over-biased. However, in this case the PMOS, threshold compensation may provide a larger drive current than the nominal value used in timing analysis. While this may cause matching problems in analog circuits, for digital circuits it actually helps setup mode timing, as long as the hold mode timing margin is not very stringent. This is usually the case for most designs and most paths even in a design with a tight hold time margin. Recall that NBTI becomes a concern because it causes timing problems by slowing down the critical paths and causing timing violations after a certain amount of stress time. For circuits like shift registers, where hold timing is critical, NBTI is most likely not an issue anyway and we simply do not deploy any compensation for them. Furthermore, Figure 10 and Figure 11 both show that even with the NBTI monitor the V_{TH} degradation is still not fully removed, so some extra amount of forward biasing still helps recover the lost margin.

If one really wants to remove the pessimism, there are at least two compatible approaches to choose from, or combine. The first is to identify the most critical paths whose timing analysis result is extremely sensitive to NBTI degradation. This can be achieved with reliability simulations, either by commercial tools supported by foundry-provided device reliability parameters (Liu & McGaughy, 2006) or implementing one's own device reliability models and incorporating them in the simulation (Wang, et al., 2007). An NBTI-aware design scheme that only upsizes those PMOS on critical paths is also presented in Yang and Saluja (2007). The difference is that, for overdesigning, accurate sizing for each transistor directly depends on modeling and simulation precision which may even reply on a correct assumption of input signal switching activity, while for monitor deployment we only need to identify the most vulnerable portions of the circuit.

The second scheme, as illustrated in Figure 12, is to only selectively turn on the monitor when NBTI degradation is large enough and a degradation constraint either on I_{DS} or V_{TH} is violated. The violation detection can be done either off-line in a special testing mode enabled at preset time points, or online with error detection structures as in the Razor architecture (Ernst, et al., 2005). Notice that when the body bias output is selected by the multiplexer, the reliability monitor should also be turned off to prevent the proxy ($P1$ in Figure 9) from being over-stressed. In addition, the pass transistor logic multiplexer needs to pass an analog output V_{OUT}, and cannot be replaced by a standard cell implementation. Pass transistor logic can be used here because V_{OUT} is connected to the body with small current and feedback; power savings can be achieved by being able to turn off the NBTI monitors.

Figure 12. Selectively power on/off the monitor to prevent over-stressing and save power (adapted from Qi & Stan, 2008)



SUMMARY

In this chapter, we proposed a threshold compensation scheme that can optimize the system energy efficiency. We demonstrated that the design can effectively mitigate the transistor threshold voltage shift due to non-ideal factors, like temperature variability and reliability degradation, and prevent the system drifting far away from its optimal operation condition. A detailed analysis of the proposed scheme shows that although ideal compensation is physically impossible, several design parameters can be tuned to maximize the compensation capability. Two example applications are demonstrated for temperature variation and NBTI-caused threshold shifts. In these examples the proposed scheme is shown to improve circuit performance, reliability and lifetime for both combinational and memory circuitry.

With little modification, the proposed scheme can be adapted for a wide range of applications. For example, compensating the threshold shift by HCI (Hot Carrier Injection) or PBTI (Positive Bias Temperature Instability) can be done very similarly as for NBTI. The scheme can also be applied to adjust shifted threshold voltages due to manufacturing process variation. Intra-die or inter-die variation can be mitigated through an on-chip proxy transistor and an off-chip reference current (I_B in Figure 1). Depending on the actual application, possible simplifications may be made, as in the case of NBTI sensor, which exploits the fact that only forward biasing is needed.

We believe adaptive and distributed computing are two viable approaches for making future computing systems more sustainable. Adaptive and distributed computing can be implemented across all the levels from the bottom circuit level to the top system level. The proposed scheme can be viewed as one such example at the circuit level. The ability to adapt to the runtime usage and environmental variations is critical to maximize the system efficiency. Distributing such compensation blocks across the whole chip can achieve more aggressive savings at a finer granularity. By adaptively compensating for both short time variation and long time shifting of transistor threshold voltages, the proposed scheme enables typical case-based design, thus effectively reducing design margins and associated chip area and power consumption. The two different example applications, i.e., compensating for short time temperature and long time NBTI aging induced threshold voltage instability, demonstrate its versatility.

REFERENCES

Agostinelli, M., Pae, S., Yang, W., Prasad, C., Kenche, D., Ramey, S., et al. (2005). Random charging effects for PMOS NBTI in ultra-small gate area devices. In *Proceedings of Annual International Reliability Physics Symposium*. Reliability Physics Symposium.

Asanovic, K., Bodik, R., Catanzaro, B. C., Gebis, J. J., & Husbands, P. K. (2006). *The Landscape of parallel computing research: A view from Berkeley*. Berkeley, CA: University of California at Berkeley.

Bellaouar, A., Fridi, A., Elmasry, M., & Itoh, K. (1998). Supply voltage scaling for temperature-insensitive CMOS circuit operation. *IEEE Transaction on Circuits and System II*, *45*(3), 415–417. doi:10.1109/82.664253

Bhardwaj, S., Wang, W., Vattikonda, R., Cao, Y., & Vrudhula, S. (2006). Predictive modeling of the NBTI effect for reliable design. In *Proceedings of IEEE Custom Integrated Circuits Conference.* IEEE Press.

Blaauw, D. T., Oh, C., Zolotov, V., & Dasgupta, A. (2003). Static electromigration analysis for on-chip signal interconnects. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *22*(1), 39–48. doi:10.1109/TCAD.2002.805728

Borkar, S. (2005). Designing reliable systems from unreliable components: The challenges of transistor variability and degradation. *IEEE Micro*, *25*(6), 10–16. doi:10.1109/MM.2005.110

Chandrakasan, A. P., Sheng, S., & Brodersen, R. W. (1992). Low-power digital CMOS design. *IEEE Journal of Solid-state Circuits*, *27*(4), 473–484. doi:10.1109/4.126534

Chen, G., Chuah, K. Y., Li, M. F., Chan, D. S., Ang, C. H., Zheng, J. Z., et al. (2003). Dynamic NBTI of PMOS transistors and its impact on device lifetime. In *Proceedings of Annual International Reliability Physics Symposium*, (pp. 196–202). Reliability Physics Symposium.

Ernst, D., Das, S., Lee, S., Blaauw, D., Austin, T., & Mudge, T. (2005). Razor: Circuit-level correction of timing errors for low-power operation. *IEEE Micro*, *24*(10), 10–20.

Hamzaoglu, F., Ye, Y., Keshavarzi, A., & Zhang, K., S., N., Borkar, S., et al. (2002). Analysis of dual-VT SRAM cells with full-swing single-ended bit line sensing for on-chip cache. *IEEE Transactions on Very Large Scale Integration (VLSI). Systems*, *10*(2), 91–95.

Hennessy, J. L., & Patterson, D. A. (2003). *Computer architecture: A quantitiative approach* (3rd ed.). London, UK: Elsevier Science Ltd.

Kang, K., Kim, K., Islam, A. E., Alam, M. A., & Roy, K. (2007). Characterization and estimation of circuit reliability degradation under NBTI using on-line IDDQ measurement. In *Proceedings of Design Automation Conference*, (pp. 358–363). Design Automation Conference.

Kang, K., Kufluoglu, H., Alam, M. A., & Roy, K. (2007). Efficient transistor-level sizing technique under temporal performance degradation due to NBTI. In *Proceedings of IEEE International Conference on Computer Design*, (pp. 934–939). IEEE Press.

Karl, E., Blaauw, D., Sylvester, D., & Mudge, T. (2006). Reliability modeling and management in dynamic microprocessor-based systems. In *Proceedings of Design Automation Conference*, (pp. 1057–1060). Design Automation Conference.

Karnik, T., Borkar, S., & De, V. (2002). Sub-90nm technologies: Challenges and opportunities for CAD. In *Proceedings of International Conference on Computer Aided Design*, (pp. 203 - 206). Conference on Computer Aided Design.

Kleissl, J., & Agarwal, Y. (2010). Cyber-physical energy systems: Focus on smart buildings. In *Proceedings of Design Automation Conference*, (pp. 749 - 754). Design Automation Conference.

Kumar, S. V., Kim, C. H., & Sapatnekar, S. S. (2006). Impact of NBTI on SRAM read stability and design for reliability. In *Proceedings of International Symposium on Quality Electronic Design.* Quality Electronic Design.

Liu, Z., & McGaughy, B. W. (2006). Design tools for reliability analysis. In *Proceedings of Design Automation Conference*, (pp. 182–187). Design Automation Conference.

Lu, Z., Lach, J., Stan, M. R., & Skadron, K. (2005). Improved thermal management with reliability banking. *IEEE Micro*, *25*(6), 40–49. doi:10.1109/MM.2005.114

Massey, J. G. (2004). NBTI: What we know and what we need to know - A tutorial addressing the current understanding and challenges for the future. In *Proceedings of the IEEE International Integrated Reliability Workshop*, (pp. 199–211). IEEE Press.

Mishra, S., Pecht, M., & Goodman, D. L. (2002). In-situ sensors for product reliability monitoring. In *Proceedings of SPIE* (pp. 10–19). SPIE. doi:10.1117/12.462807

Mitra, S. (2007). Circuit failure prediction enables robust system design resilient to aging and wearout. In *Proceedings of the IEEE International On-Line Testing Symposium*, (p. 123). IEEE Press.

Mostafa, H., Anis, M., & Elmasry, M. (2011). Adaptive body bias for reducing the impacts of NBTI and process variations on 6T SRAM cells. *IEEE Transactions on Circuits and Systems. I, Regular Papers*, *58*(12), 2859–2871. doi:10.1109/TCSI.2011.2158708

Park, C., John, J., Klein, K., Teplik, J., Caravella, J., Whitfield, J., et al. (1995). Reversal of temperature dependence of integrated circuits operating at very low voltages. In *Proceedings of the International Electron Devices Meeting*, (pp. 71 - 74). Electron Devices Meeting.

Paul, B. C., Kang, K., Kufluoglu, H., Alam, M. A., & Roy, K. (2006). Temporal performance degradation under NBTI: Estimation and design for improved reliability of nanoscale circuits. In *Proceedings of European Design and Test Conference*, (pp. 780–785). European Design and Test Conference.

Pedram, M. (2009). Green computing: Reducing energy cost and carbon footprint of information processing systems. In *Proceedings of the ACM Great Lakes symposium on VLSI.* ACM Press.

PTM. (2011). *PTM technology*. Retrieved from http://ptm.asu.edu/.

Qi, Z., Meyer, B., Huang, W., Ribando, R., Skadron, K., & Stan, M. (2010). Temperature-to-power mapping. In *Proceedings of IEEE International Conference on Computer Design*, (pp. 384 - 389). IEEE Press.

Qi, Z., & Stan, M. R. (2007). Accurate back-of-the-envelope transistor model for deep sub-micron MOS. In *Proceedings International Conference on Microelectronic Systems Education.* Microelectronic Systems Education.

Qi, Z., & Stan, M. R. (2008). NBTI resilient circuits using adaptive body biasing. In *Proceedings of the ACM Great Lakes symposium on VLSI*, (pp. 285-290). ACM Press.

Sakurai, T., & Newton, A. R. (1991). A simple MOSFET model for circuit analysis. *IEEE Journal of Solid-state Circuits*, *38*(4), 887–893.

Saluja, K., Vijayakumar, S., Sootkaneung, W., & Yang, X. (2008). NBTI degradation: A problem or a scare? In *Proceedings of the VLSI Design, 2008: VLSID 2008 21st International Conference*, (pp. 137 - 142). VLSI.

Shakeri, K., & Meindl, J. D. (2003). Temperature variable supply voltage for power reduction. In *Proceedings of IEEE Computer Society Annual Symposium on VLSI.* IEEE Press.

Stan, M., Cabe, A., Ghosh, S., & Qi, Z. (2007). Teaching top-down ASIC/SoC design vs bottom-up custom VLSI. In *Proceedings of the International Conference on Microelectronic Systems Education (MSE).* MSE.

Tang, S., Narendra, S., & De, V. (2003). Temperature and process invariant MOS-based reference current generation circuits for sub-1V operation. In *Proceedings of of International Symposium on Low Power Electronics and Design.* Low Power Electronics and Design.

Tschanz, J. W., Kao, J. T., Narendra, S. G., & Nair, R. A. A., D., Chandrakasan, A. P., et al. (2002). Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage. *IEEE Journal of Solid-state Circuits*, *37*(11), 1396–1402. doi:10.1109/JSSC.2002.803949

Vattikonda, R., Wang, W., & Cao, Y. (2006). Modeling and minimization of PMOS NBTI effect for robust nanometer design. In *Proceedings of Design Automation Conference*, (pp. 1047–1052). Design Automation Conference.

Wang, W., Yang, S., Bhardwaj, S., Vattikonda, R., Vrudula, S., Liu, F., et al. (2007). The impact of NBTI on the performance of combinational and sequential circuits. In *Proceedings of Design Automation Conference*, (pp. 364-369). Design Automation Conference.

Yang, X., & Saluja, K. (2007). Combating NBTI degradation via gate sizing. In *Proceedings of International Symposium on Quality Electronic Design*, (pp. 47–52). Quality Electronic Design.

Zhang, Y., & Stan, M. (2007). Temperature-aware circuit design using adaptive body biasing. In *Proceedings of the ACM Great Lakes Symposium on VLSI*, (pp. 84–89). ACM Press.

Zhao, W., & Cao, Y. (2006). New generation of predictive technology model for sub-45nm early design exploration. *IEEE Transactions on Electron Devices*, *53*(11), 2816–2823. doi:10.1109/TED.2006.884077

Zhou, Q., & Mohanram, K. (2006). Gate sizing to radiation harden combinational logic. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *25*(1), 155–166. doi:10.1109/TCAD.2005.853696

ADDITIONAL READING

Basoglu, M., Orshansky, M., & Erez, M. (2010). NBTI-aware DVFS: A new approach to saving energy and increasing processor lifetime. In *Proceedings of the 16th ACM/IEEE International Symposium on Low Power Electronics and Design.* ACM Press.

Bsoul, A. A., Manjikian, N., & Shang, L. (2010). Reliability and process variation-aware placement for FPGAs. In *Proceedings of the Conference on Design, Automation and Test in Europe.* Design Automation and Test.

Dadgour, H., & Banerjee, K. (2010). Aging-resilient design of pipelined architectures using novel detection and correction circuits. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE 2010).* Design Automation and Test.

Firouzi, F., Kiamehr, S., & Tahoori, M. B. (2011). A linear programming approach for minimum NBTI vector selection. In *Proceedings of the Great Lakes Symposium on VLSI.* ACM Press.

Fu, X., Li, T., & Fortes, J. (2008). NBTI tolerant microarchitecture design in the presence of process variation. In *Proceedings of the Microarchitecture IEEE/ACM International Symposium*, (pp. 399-410). IEEE Press.

Guardiani, C., Shibkov, A., Brambilla, A., Gajani, G., Appello, D., Piazza, F., et al. (2009). An I-IP based approach for the monitoring of NBTI effects in SoCs. In *Proceedings of the IEEE International On-Line Testing Symposium.* IEEE Press.

Kothawade, S., Chakraborty, K., & Roy, S. (2011). Analysis and mitigation of NBTI aging in register file: An end-to-end approach. In *Proceedings of the International Symposium on Quality Electronic Design (ISQED).* ISQED.

Rabaey, J., Chandrakasan, A., & Nikolic, B. (2003). *Digital integrated circuits: A design perspective* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Rahaman, M. S., Duan, Q., & Chowdhury, M. H. (2011). Spatial- and temporal-reliability aware design for nano-scale VLSI circuits. In *Proceedings of the IEEE International Symposium on Circuits and Systems.* IEEE Press.

Rakesh, R., & Yadav, K. S., & Jaipal. (2011). Reduction of leakage by input vectors with constrained NBTI degradation. *International Journal of Information Technology and Knowledge Management*, *4*(1), 163–167.

Ramakrishnan, K., Wu, X., Vijaykrishnan, N., & Xie, Y. (2008). Comparative analysis of NBTI effects on low power and high performance flip-flops. In *Proceedings of the IEEE International Conference on Computer Design*, (pp. 200 - 205). IEEE Press.

Salfelder, F., & Hedrich, L. (2011). An NBTI model for efficient transient simulation of analogue circuits. In *Proceedings of edaWorkshop*. EDA.

Siddiqua, S. G. T. (2009). NBTI-aware dynamic instruction scheduling. In *Proceedings of the IEEE Workshop on Silicon Errors in Logic - System Effects*. IEEE Press.

Siddiqua, T., & Gurumurthi, S. (2009). NBTI-aware dynamic instruction scheduling. In *Proceedings of the IEEE Workshop on Silicon Errors in Logic - System Effects*. IEEE Press.

Song, J., Han, Y., Zhang, L., Li, H., Li, X., & Yan, G. (2009). M-IVC: Using multiple input vectors to minimize aging-induced delay. In *Proceedings of the Asian Test Symposium.* Asian Test.

Subramanyan, P., Jangir, R. R., Tudu, J., Larsson, E., & Singh, V. (2009). Generation of minimal leakage input vectors with constrained NBTI degradation. In *Proceedings of the 8th East-West Design and Test Workshop.* East-West Design and Test.

Vazquez, J., Champac, V., Ziesemer, A., Reis, R., Teixeira, I., Santos, M., et al. (2009). Built-in aging monitoring for safety-critical applications. In *Proceedings of the 15th IEEE International On-Line Testing Symposium.* IEEE Press.

Vazquez, J., Champac, V., Ziesemer, A., Reis, R., Teixeira, I., Santos, M., et al. (2010). Low-sensitivity to process variations aging sensor for automotive safety-critical applications. In *Proceedings of the VLSI Test Symposium (VTS).* VTS.

KEY TERMS AND DEFINITIONS

Design Margin/Guardband: The extra capability designed into a circuit system slightly beyond the minimum requirement from design specification in terms of performance, reliability, power, etc, to compensate for all kinds of variations or uncertainties. This can translate to extra circuit area, extra power consumption, etc.

Feedback: Used in the context of control theory, meaning the system output is fed back into as part of the input to control the system. Feedback improves stability of a system.

Hot Carrier Injection (HCI): A phenomenon in which high energy ("hot") electrons tunnel through the gate dioxide of a transistor, causing threshold voltage drift and degraded transistor performance.

NBTI/PBTI: Negative/Positive Bias Temperature Instability. NBTI happens in PMOS with negative voltage bias between gate and source terminal. The effect is increased threshold voltage and larger transimpedance. The degradation increases with the biasing time and elevated temperature, but can partially recover when the bias is removed. PBTI is the symmetric mechanism for NMOS.

Open/Closed Loop: Systems with/without feedback mechanism

Reliability Degradation/Aging: The performance of a transistor can change (usually resulting in inferior performance, e.g., slower transition, reduced noise margin, or even failure to turn on) after it is turned on and passes currents for some time. Mechanisms include NBTI (Negative Bias Temperature Instability), HCI (Hot Carrier Injection), TDDB (Time Dependent Dioxide Breakdown), etc.

Threshold Voltage: The gate voltage needed for forming a low-resistance channel (inversion layer) between the source and drain of a transistor. In oversimplified terms, it can be considered a point beyond which the transistor is "turned on" (passing current) and below which "turned off" (open circuit).

Transimpedance/Transconductance: For a two-port circuit, transimpedance is the ratio of the output voltage change to the input current change, and transconductance is the ratio of the output current change to the input voltage change.

Transistor/NMOS/PMOS: A transistor (also called MOS or MOSFET) is the basic building element for modern integrated circuits. There are two types of transistors: NMOS (N-type MOS transistor, also called NFET) and PMOS (P-type MOS transistor, also called PFET).

Variation: Circuit uncertainty and drift from the values designed for in all aspects of design specifications. Variation can come from many sources including manufacturing process (process variation), power supply instability (power variation), circuit operating temperature (thermal variation), aging degradation (aging variation), etc.

This work was previously published in Sustainable ICTs and Management Systems for Green Computing, edited by Wen-Chen Hu and Naima Kaabouch, pp. 201-230, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 19
Modeling Closed Loop Supply Chain Systems

Roberto Poles
University of Melbourne, Australia

ABSTRACT

In the past, many companies were concerned with managing activities primarily along the traditional supply chain to optimize operational processes and thereby economic benefits, without considering new economic or environmental opportunities in relation to the reverse supply chain and the use of used or reclaimed products. In contrast, companies are now showing increased interest in reverse logistics and closed loop supply chains (CLSCs) and their economic benefits and environmental impacts. In this chapter, our focus is the study of remanufacturing activity, which is one of the main recovery methods applied to closed loop supply chains. Specifically, the authors investigate and evaluate strategies for effective management of inventory control and production planning of a remanufacturing system. To pursue this objective, they model a production and inventory system for remanufacturing using the System Dynamics (SD) simulation modeling approach. The authors primary interest is in the returns process of such a system. Case studies will be referred to in this chapter to support some of the findings and to further validate the developed model.

INTRODUCTION

Industry in general, and society more broadly, have come to recognize the limited availability of natural resources and are moving towards the manufacture of more environmentally friendly products and the recovery of resources. For this reason, the modern trend, particularly in developed countries, is to use fewer environmental resources such as water, air and raw materials to manufacture products. Moreover, interest in strategic sustainability is growing among multinational companies, some of whom are developing sustainability reports to demonstrate both their concern for the environment and their commitment to conducting socio-ecological activities in business. In addi-

DOI: 10.4018/978-1-4666-1945-6.ch019

tion, sustainability can be used as a competitive strategy to create company branding, comply with government regulations on the environment and optimize the cost of operational processes.

This emerging economic and environmental consciousness within business has increased the focus on reverse logistics activity (company processes that recapture value from product returns) over the last decade (Blumberg, 2005). Indeed, this reverse logistics activity, particularly remanufacturing (the process of reusing returned products in production), can play an important role in sustainability, as well as in developing competitive strategies aimed at reducing the use of natural resources and recovering value from used products. However, several factors make the development of reverse logistics processes difficult. In particular, the complex integration between the forward (from the producer to the consumer) and the reverse (from the consumer to the producer) supply chains can negatively affect operations and logistics management activities such as production planning, inventory control and distribution planning.

According to the Reverse Logistics Executive Council, reverse logistics is the process of moving goods from the point of consumption to the point of origin for the purpose of either recapturing value or proper disposal. Stock (2001) has defined reverse logistics as "the term most often used to refer to the role of logistics in product returns, source reduction, recycling, material substitution, reuse of material, waste disposal, and refurbishing, repair and remanufacturing" (p. 5).

Companies can pursue several methods of recapturing value from returns, and carry out a range of recovery methods (Kulwiec, 2006). For example, products can be reused directly after cleaning or reconstruction. This is a common practice for items such as used pallets, bottles/glass or containers. However, products whose parts or materials need to be repaired or replaced can be reused after repair as rebuilt or used products. Another method of recovery is remanufacturing, a process in which parts and materials from returned products are reused for production. Remanufacturing requires more extensive work, since the returned product must be completely disassembled, its parts and modules examined and either repaired or replaced, and then reassembled into a new product. Remanufacturing is practiced in many industries, including for photocopiers, computers, telecommunications equipment, automotive parts, office furniture and tires. A final recovery option is recycling. In this case, some or all of the parts and materials from returned products can be processed to make different products (Kulwiec, 2006).

The concept of reverse logistics has changed in recent years (Dekker, Fleischmann, Inderfurth, & Wassenhove, 2004). More specifically, the concept of a closed loop supply chain (CLSC) has been developed to refer to the complete loop from the customer, back to the plant, through a reprocessing operation, and then back to the customer (French & LaForge, 2006). Closed loops consist of two integrated supply chains—a forward and a reverse chain—through which a recovered product re-enters the original forward chain (Wells & Seitz, 2005). CLSC is sometimes treated as an extension of the traditional concept of supply chain management. In this case, reverse logistics is not managed independently of forward logistics, but rather both processes form part of a complete supply chain process whereby products start with the manufacturer to reach the customer, then come back to the plant and return to the customer once more. This process has been defined by this new concept of the closed loop supply chain (Eoksu, Sungwon, Haejoong, & Jinwoo, 2004).

The integration of reverse logistics activities within the structure of the original production and distribution systems leads to additional complexity in the closed loop supply chain system. This complexity, which can hinder the integration process, comes from the significant differences between the forward and reverse supply chains. The latter are characterized by operational and business factors that disrupt the traditional approach to dealing with supply chain management, such as: (1) the complexity of forecasting returns; (2) a lack of uniformity in returns quality; (3) reverse logistics costs being less directly visible; and (4) production and inventory management being affected by the flow of returns. These factors make it necessary to rethink and re-plan the original supply chain. In particular, in the context of the remanufacturing process, production and inventory management entail the added complexity of coordination between the remanufacturing and production activity in terms of size of orders and lead times. Moreover, such coordination must account for the new kind of inventory generated by the additional flow of collected returns to be integrated into the remanufacturing process. However, this additional flow is not directly available to the manufacturer because of the unpredictability of the quantity, timing and quality of the products returned by customers.

The uncertainty around the quantity and timing of returns is one of the main factors that make the implementation of closed loop supply chain processes difficult, particularly in the case of integration between the forward and reverse supply chains. For example, the difficulty of determining the quantity of used products that will be returned by customers negatively affects remanufacturing and traditional production planning. Moreover, the lack of tools and guidelines on planning, controlling and managing remanufacturing operations has limited the growth of the remanufacturing sector (Guide, 2000). If not well designed, closed loop supply chain activities such as remanufacturing and disposal can increase company costs (Inderfurth, 2005). For this reason, a company objective is to integrate the reverse and forward supply chains so as to minimize the total cost and consequently obtain economic benefits.

Mindful that these problems add complexity to this field of research, in this chapter we aim primarily to model the factors affecting a production and inventory system that combines returns and remanufacturing, and to evaluate effective control strategies that address dynamic production and inventory management issues in order to improve the performance of the system. For this purpose, the System Dynamics (SD) simulation modeling approach (Forrester, 1958, 1961)—a methodology used for studying and managing complex feedback systems, particularly business and social systems—is adopted to model a production and inventory system for remanufacturing within the context of closed loop supply chains in order to understand the complex and dynamic interaction of factors that affect the behavior of the system. Our primary interest is the returns process in the context of closed loop supply chains. Moreover, this model is used to evaluate strategies for the effective management of remanufacturing and returns processes.

In order to assess some of the findings and to further validate the developed model, the methodological approach taken in this chapter is based on case study research. The selection of companies employed as case studies was based on their engagement in remanufacturing and returns processes, which made them useful for our study. The companies studied are: (1) the Australian Mobile Telecommunications Association (AMTA), which has commenced a national recycling program for mobile phones; and (2) Fuji Xerox Australia, which is involved in the remanufacturing of assemblies and sub-assemblies of printers and copiers.

BACKGROUND

The introduction of new environmental legislation in recent years has strengthened the need to focus on reverse logistics and CLSC among logistics operators. While legislation introduced in Europe, North America and Japan encourages this awareness, many corporations have proactively taken measures in anticipation of evolving environmental performance requirements (Savaskan, Bhattacharya, & Wassenhove, 2004). Several examples of government regulations can be considered. In Japan any products purchased by the government must contain recycled materials and the European Union (EU) has issued a directive for producer responsibility to collect, process and recycle waste from both "white goods" (e.g. refrigerators, washing machines and freezers) and "brown goods" (e.g. TVs and speakers) (Kulwiec, 2006). In Europe, many countries have forced industry to develop collection and recycling systems in order to reduce waste. This has led to the issuing of several environmental regulations such as the EU Directive 2000/53/EC (Blanc, Fleuren, & Krikke, 2004) for responsibility in taking care of used products, the EU Directive 2002/96/EC and 2003/108/EC for electrical and electronic equipment, and the EU Directive 2002/525/EC for end-of-life vehicles (Tang, Grubbström, & Zanoni, 2007). According to the US Environmental Protection Agency (EPA), in the United States (US) the amount of waste generated has increased from 88 million tons in the 1960s to 196 million tons in 1990 alone. Consequently, companies need to develop techniques for product recovery and waste management (Gungor & Grupta, 1999). Furthermore, hundreds of environmental laws and regulations for recycling operations and responsibility for packaging recovery have been developed (Kulwiec, 2006). Local governments in North America promote the reduction of landfill use, which is the main driver of non-toxic solid waste, pushing companies towards improving and innovating in their manufacturing activities (Biehl, Prater, & Realff, 2007). This places an onus on manufacturing firms to use reverse logistics activities as a form of extended producer responsibility (EPR), which makes them responsible for their products throughout their life cycles (Klausner & Hendrickson, 2000).

In the US, the precise extent of reverse logistics activities is difficult to determine because most companies have not quantified them (Reverse Logistics Executive Council, 2007). However, reverse logistics activities account for a significant portion of logistics costs in the US. These are estimated to account for approximately 10.7% of the US economy and approximately 4% of the total logistics costs (Reverse Logistics Executive Council, 2007). Reverse logistics costs are estimated to be approximately 0.5% of total US gross domestic product (GDP), which equated to around US$58.34 billion in 2004. In 1999 the total value of returned merchandise in the US, with an estimated handling cost of $40 billion, was $62 billion (ReturnBuy, 2000). The US is not the only country where reverse logistics activities are increasing. In Europe, the annual production of remanufactured automotive parts was approximately 20,000,000 units in 2005, and is expected to increase to 30,000,000 units in 2015 (Automotive Parts Rebuilders Association, 2007). In Australia, the Eco Manufacturing Centre of Fuji Xerox, a company that continually adopts new remanufacturing programs, has aimed for zero waste discharge and has achieved 90% reuse and recycling (Environment Protection Authority, 2002).

Based on these data, it is evident that many companies have realized that reverse logistics/CLSC, particularly remanufacturing, is an important competitive and strategic component of their business mission. Indeed, the use of reverse logistics in the business sector is increasing not only because of the implementation of more stringent environmental regulations, but also for reasons related to competition. The results of a survey involving 1,200 logistics managers and more than 150 managers with reverse logistics responsibilities in the US found that 65% of companies believe that returns management is an important strategic tool for their business (Rogers & Tibben-Lembke, 2001). Indeed, through reverse logistics and closed loop supply chain activities companies can fulfill the environmental responsibilities stipulated by government regulations and at the same time optimize operational processes. Such optimization can be achieved both by reducing operating expenses, and by improving company image through a reduced use of resources. The cost of remanufacturing, for example, is typically 40–60% of the cost of manufacturing, and remanufactured products are of the same quality as new products, and sold with the same warranties (Mitra, 2007). Moreover, customers have become more environmentally conscious, which in recent years has led to an increase to over US$2,000 billion worth of environmentally friendly products on the global market (Mitra, 2007). This growing pressure to improve the market competitiveness of companies through the reverse logistics process has pushed researchers to analyze, model and explain why and how reverse logistics can lead to economic and environmental benefits. A number of studies describe the role of reverse logistics in economic and environmental activities during the product development process. In particular, several models have been developed to support managerial decision-making and to optimize processes in different reverse logistics areas. The scope of these models is mostly to minimize costs and optimize profits through analysis of the parameters and variables as defined in the modeling method.

However, the characteristic of variability in the quality, quantity and timing of returned products and the integration of the returns flow within the original forward supply chain make reverse logistics activities, particularly remanufacturing, difficult to plan, control and manage (Guide & Wassenhove, 2001). For this reason, systems that are not appropriate for dealing with returns could increase operating expenses. In a study by Guide (2000), 61.5% of the firms under study were found to have no control over the timing or quantity of returns. He discusses several characteristics of recoverable manufacturing systems that complicate production planning and points out that there is a significant lack of specific technologies and techniques for remanufacturing logistics. These include: uncertainty around the timing and quantity of returns; balancing returns with demands; disassembly of returned products; and materials recovery uncertainty, which are all factors that require considerable research. These characteristics could be addressed by focusing on several issues, such as the methods (e.g. leasing, deposits) used to reduce uncertainty around timing and quantity of returns; forecasting models; aggregate production planning models that consider returns; and models that support material recovery planning and prediction based on the age and usage rate of products.

Production and inventory management requires appropriate control mechanisms to integrate the return flow of used products within the material planning for the forward flow (Fleischmann et al., 1997). This can vary for different reverse logistics situations. For example, for companies whose business is recycling, returns are the only inventory resources for the forward production process and used products or materials are the only raw materials. Traditional inventory control methods might be satisfactory in this context. However, the mechanism is different for remanufacturing or reuse where used products are returned for introduction into the main production stream. In this case, returned goods consist of an additional inventory source to the usual inventory procured from outside. Moreover, this additional flow is not directly available to the manufacturer because of the unpredictable factors of quantity, time and quality of the products returned. Hence, inventory management can be made particularly complicated by remanufacturing activities since key information such as that related to on-hand inventory, lead time and yield is not clearly known (Toktay, Wein, & Zenios, 2000).

For these reasons, in this study we model a production and inventory system in which production is integrated with remanufacturing activity. Several authors have conducted research into such a system. They have focused mainly on: production and remanufacturing lead times (Inderfurth & van der Laan, 2001; Kiesmuller, 2003; Kiesmuller & Minner, 2003; van der Laan, Salomon, & Dekker, 1999); optimization procedures for inventory levels and economic order quantity (Kiesmuller & van der Laan, 2001; Koh, Hwang, Sohn, & Ko, 2002; Teunter, 2001; van der Laan, Dekker, & Salomon, 1996; van der Laan, Dekker, Salomon, & Ridder, 1996); comparisons between pull and push strategies (van der Laan & Salomon, 1997; van der Laan, Salomon, Dekker, & Wassenhove, 1999; van der Laan & Teunter, 2006); and capacity planning (Georgiadis, Vlachos, & Tagaras, 2006; Kleber, 2006; Vlachos, Georgiadis, & Iakovou, 2007). However, to our knowledge no previous study has examined a system in which the returns process is modeled and analyzed in relation to the particular system variables involved in our research. Several inventory models for the remanufacturing process have been developed where returns are included as exogenous variables, without any or with only simple correlation between demand and returns. Many of these have adopted simple assumptions regarding the returns process such as the homogeneous Poisson Process for demand and/or return flow, or that returns are independent of the demand (de Brito & Dekker, 2003).

MODEL BUILDING

The modeling process undertaken in this research was characterized by a sequence of iterative activities and stages that involved continuous revisions and changes. Indeed, this modeling can be defined as a continual process of iteration among problem articulation, the generation of hypotheses, data collection, model formulation, testing and analysis (Sterman, 2000). In this section, the final results of this process of iteration, which led to the model building for the production and inventory system for remanufacturing, are presented.

Description of the System under Study

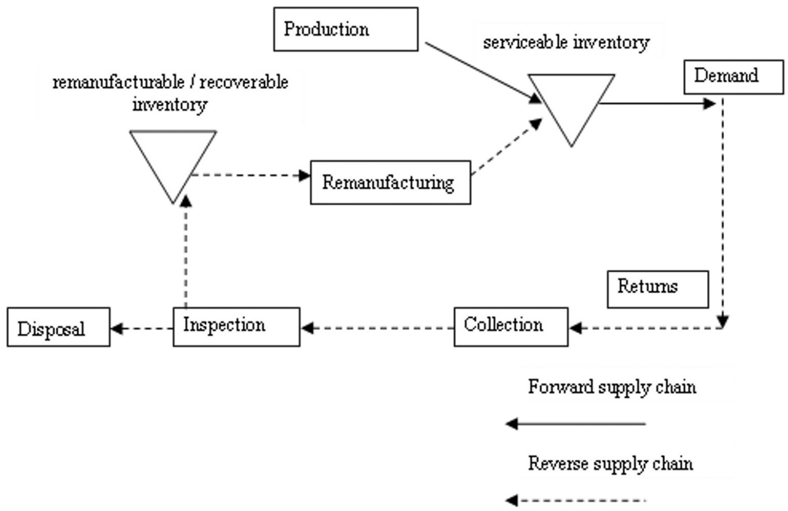
The purpose of this chapter is to model the factors affecting a production and inventory system that combines returns and remanufacturing within the context of closed loop supply chains, and to evaluate effective control strategies aimed at improving the performance of the system, focusing particularly on the returns process. For this purpose, we model and simulate a single-product production and inventory system for remanufacturing within the context of closed loop supply chains. The system involves several operations including: production of new products; collection and inspection of used products; remanufacturing; and disposal. Our focus in this study is on the return of products by customers/product users at the end of the products' useful life; other returns such as product recalls and B2B commercial returns are excluded from examination.

The system under consideration is depicted in Figure 1. The forward supply chain involves the production of new products to meet customer demand. After product use, returns are collected, inspected and either stored as remanufacturable/recoverable inventory or disposed of depending on whether the quality of returns is deemed suitable for remanufacturing according to the company's policy on quality standards. The serviceable inventory, used to fulfill external demands, is fed by the production of new or remanufactured products which are as good as new. Production and remanufacturing activities are important components of any production and inventory system for remanufacturing. Equally important to the process is analysis and decision-making regarding inventory, operational and marketing activities.

The following assumptions are adopted throughout this analysis in order to simplify the system and facilitate the modeling process by focusing on the most important factors:

- Uncontrollable disposal is not considered, whereby instead of the product returning to the remanufacturer, it is disposed of in an uncontrolled manner, sometimes in opposition to the manufacturer's instructions or environmental regulations.

Figure 1. Production and inventory system for remanufacturing



- The planned disposal of recoverable inventory is not considered.
- The capacity of several activities such as collection, inspection, remanufacturing and production are considered infinite.
- Backordering and lead times are not considered.

In our model, the returns rate incorporates the uncertainty around the quantity and timing of returns and a pull inventory control policy is applied. This policy is implemented through reorder point inventory replenishment policies, which are basic features of several industries in the context of supply chain/inventory planning. The returns rate, which is used to calculate the number of returns after the time of use, is represented as the ratio between the probable returns flow of sold products and the demand. The probable returns flow and the time of use are calculated on the basis of the relationship between two factors: the *return index* and the *residence time*. The latter is the factor defined in the study of Georgiadis et al. (2006), and represents the average time for which a product stays with its customer before it is returned.

The ways in which companies manage the returns process for products sold to customers, through service agreements and sales contracts with retailers or the customers themselves, can influence the returns rate and particularly the quantity of returns. For example, leasing contracts ensure that almost all products are returned after the residence time. In our model, the *service agreement with customer* factor is used to relate the quantity of returns with the demand for different products in different industries. *Customer behavior* is another factor that can influence this relationship. The attitude of the customer in terms of their return activity and their response to a company's returns process incentives can affect the returns rate and in particular the likelihood of a particular product being returned. Hence, the *return index* is obtained by considering the relationship between company incentives/service agreements aimed at recovering used products and actual customer behavior in returning products. Put simply, the return index is the tendency of the product to be returned by the customer during its lifetime, which varies across different products and industries. This is explained in more detail in the next section.

Quantitative Modeling

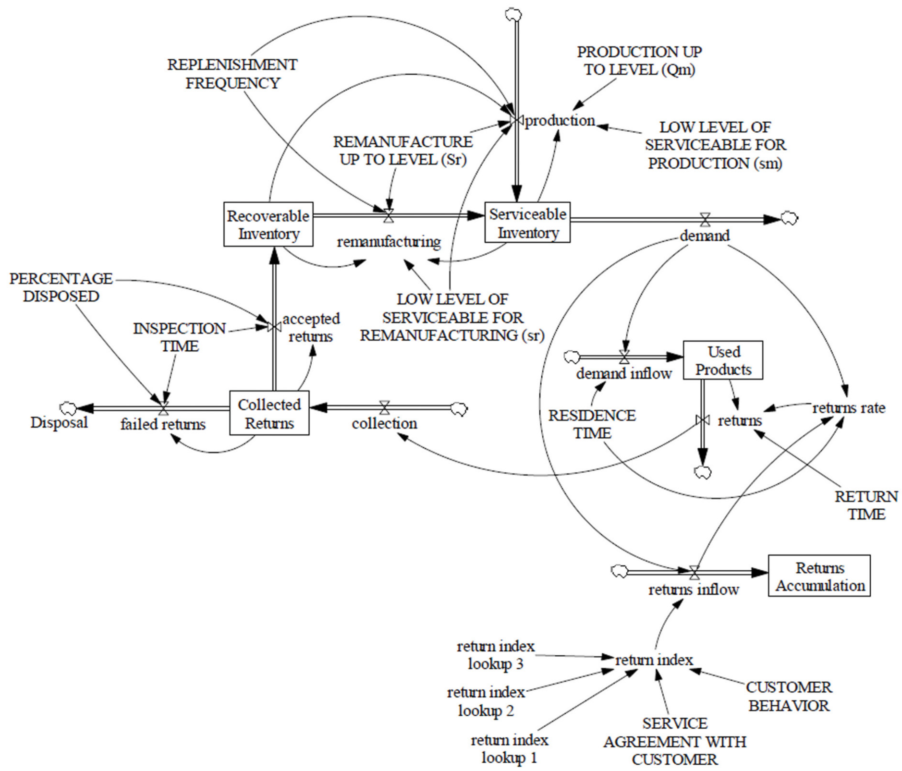
A System Dynamics (SD) simulation modeling approach is adopted in this research. SD was introduced in the early 1960s by Jay Forrester (1958, 1961) as a modeling and simulation methodology aimed at dealing with the dynamics and controllability of management systems (Coyle, 1996). The SD method is commonly used to analyze how the dynamic behavior patterns of system variables change in response to dynamic inputs. Controllability refers to the "control systems" (Coyle, 1996) by which the policies employed and applied in the system structure control system behavior over time. The objective of using SD is to identify strategies to improve system performance (Sterman, 2000). For these reasons, SD has become a computer-aided method of analyzing and solving complex problems, particularly in the area of policy analysis and design. It is applied in a number of fields, including: corporate planning and policy design; economic behavior; public management; biological and medical modeling; energy and environmental studies; social science; dynamic decision-making; complex nonlinear dynamics; software engineering; and supply chain management (Angerhofer & Angelides, 2000). The choice of a simulation approach, in particular the SD approach, rather than other methods such as analytical approaches was based on the recognition that SD can model an entire system in which several policies and factors can be used to evaluate strategies aimed at improving system performance. Moreover, SD can handle the issues arising from models in which dynamic forces and nonlinear relationships play a significant role.

All business and social systems contain a host of different asset stocks or accumulation of resources which change according to their physical inflows and outflows (Morecroft, 2007). According to Morecroft, this stock and flow structure of systems reveals the operations behind the causal influences or relationships among system variables. In SD, the stock and flow structure of systems is represented by the Stock and Flow Diagram (SFD). Such a diagram is obtained by identifying the stock, flow and auxiliary variables among the variables used to model a system. The stock variables determine the state of the system and are represented by an accumulation of the difference between the inflow to a process and its outflow (Sterman, 2000), while the flow variables determine the physical flows in the system and generate change in the stocks, which is then used to make decisions. The auxiliary variables can be useful for clarifying the structure and process of the model. They represent constants or exogenous inputs into the model, as well as converters or intermediate variables used in the mathematical equations of the model (Kirkwood, 1998; Sterman, 2000). The objective in developing an SFD is to analyze and define the dynamic relationships among stock, flow and auxiliary variables through mathematical equations in order to run simulations of the model.

The SFD for the production and inventory system for remanufacturing is presented in Figure 2. The rectangles represent stock variables which equate with accumulations of items, while the valves represent flow variables which correspond to the physical flow of items feeding or depleting the stocks. The physical flow of items is represented by a double line with arrows, while the flow of information (i.e. connections among variables and their relationships used for mathematical formulations) is represented by a single line with arrows. The auxiliary variables shown in upper case letters represent constants or exogenous inputs, while those in lower case letters represent the converters used in calculations.

In the diagram above, an increase/decrease in *returns* increases/decreases the rate of *collection*, which in turn increases/decreases the level of *Collected Returns*. At this stage of the process, returned products are inspected in order to check their quality and remanufacturability. Failed items decrease the level of *Collected Returns*, through a *failed returns* flow, and at the same time increase

Figure 2. Stock and flow diagram



the level of *Disposal* which represents the quantity of non-reusable items that are disposed of. The flow rate of failed items depends on the value of *percentage disposED* and *inspection time*. The former affects the flow rate as any change in the percentage leads to a change in the quantity of failures. For example, an increase in the percentage value of disposal leads to an increase in the flow of failed items for a given time period. *Percentage disposED* also represents the quality standards policy of the company and is affected by several parameters and techniques used to check returned items. It is defined as an average percentage of collected returns that are disposed of, and differs for different products and different quality standard policies adopted (Vlachos et al., 2007). *Inspection time* represents the period of time required to inspect collected items. This affects the inspection flow, as a faster/slower inspection time leads to an increase/decrease in the flow.

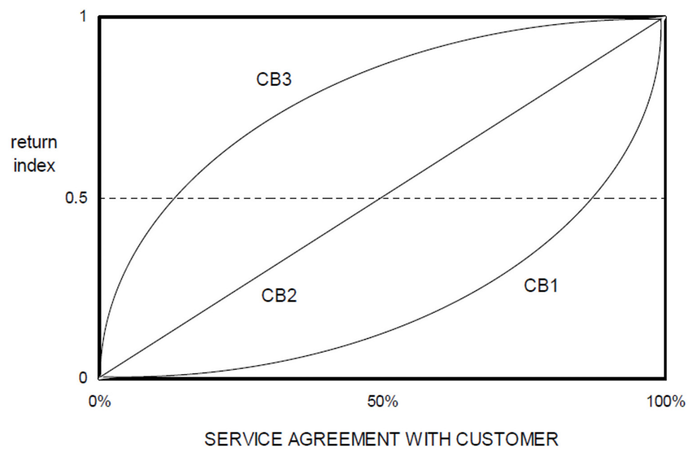
Accepted items increase the level of *Recoverable Inventory* that is ready to be remanufactured through the *accepted returns* flow. The flow rate of accepted items depends inversely on the value of *percentage disposED*, as a lower percentage of disposed items leads to a higher number of remanufacturable items.

Remanufacturable items are stored as *Recoverable Inventory* from which items are used for remanufacturing purposes when necessary and stored as *Serviceable Inventory* in order to fulfill customer *demand*. In this system, remanufacturing, which is preferred to the more expensive production activity, occurs when necessary as a pull inventory policy is applied. Several studies have previously modeled push and pull inventory policies in a production and inventory system for remanufacturing (Kiesmuller, 2003; van der Laan & Salomon, 1997; van der Laan, Salomon, & Dekker, 1999). A more detailed picture of the inventory pull policy was provided in a study by van der Laan, Salomon and Dekker (1999), who similarly did not consider the disposal of recoverable inventory in their model. According to these authors, *REMANUFACTURE UP TO LEVEL* (*Sr*) and *LOW LEVEL OF SERVICEABLE FOR REMANUFACTURING* (*sr*) are two variables that affect the *remanufacturing* rate and are used to implement a pull policy in the system. *Sr* represents the upper value limit for remanufactured batches, while *sr* represents the lower value for remanufactured batches as well as the level of a *Serviceable Inventory* at which a remanufacturing batch is required. *Sr* - *sr* represents the level of *Recoverable Inventory* for which it is possible to produce a remanufacturing batch. The pull policy is represented by the *Recoverable Inventory* level. Only when this level reaches the difference (*Sr* - *sr*), and the *Serviceable Inventory* level is lower than *sr*, is it possible to produce a remanufacturing batch. Two additional variables that affect production flow are incorporated to implement the pull inventory policy: *PRODUCTION UP TO LEVEL* (*Qm*), which is the upper value for production batches; and *LOW LEVEL OF SERVICEABLE FOR PRODUCTION* (*sm*). Production activity is only used to increase the *Serviceable Inventory* level when the *Recoverable Inventory* level is lower than (*Sr* - *sr*) and the *Serviceable Inventory* level is lower than *sm* (the *Serviceable Inventory* level at which a production batch is required). However, production flow is mainly affected by *Sr* and *sr*, because it is only when the *Recoverable Inventory* level is lower than *Sr* - *sr* and the *Serviceable Inventory* level reaches *sm* that a production batch is manufactured and stored in *Serviceable Inventory*. This strategy increases the cost of the *Recoverable Inventory* but reduces the cost of the *Serviceable Inventory* which is usually more expensive. Moreover, remanufacturing is preferred to production activity, as *sm* is lower than *sr*. *Remanufacturing* flow is also affected by the *replenishment frequency* of the inventory. *Replenishment frequency* represents the time taken to replenish remanufacturing orders and an increase/decrease in its value generally leads to a decrease/increase in the order size.

The returns process starts with customer demand, which depletes the *Serviceable Inventory* level. Product demand is considered equal to sales. After a period of time or *RESIDENCE TIME*, products in use can be considered to be used products. This is represented by the flow between the rate of the variable *demand inflow* and the level of *Used Products*. The variability of *RESIDENCE TIME* represents the uncertainty that affects the timing of returns in a closed loop supply chain. For this reason, in this model not all used products are considered to be returns but rather as possible returns after an average period of use that is dependent on the type of product. A portion of these used products become *returns*, which are consequently collected. This is represented by the physical flow in which *returns* deplete the *Used Products* level and the information flow between *returns* and *collection*.

Uncertainty around the quantity of used products returned by customers negatively affects collection, remanufacturing, production planning and inventory control. For this reason, several variables are used to reduce the effect of this uncertainty and set the quantity of returns. The *return index* is used to set the number of returns based on customer demand. The number of returns is represented by the *returns inflow*, which is influenced by both *returns index* and *demand*. Two parameters influence the *return index*: *SERVICE AGREEMENT WITH CUSTOMER* and *CUSTOMER BEHAVIOR*. The former defines the level of the service agreement or incentives that the company offers to the customer at the end or during the use of the product in order to stimulate the returns process. However, it also represents the level of responsibility the company has towards the recovery of its own products. The latter parameter, *CUSTOMER BEHAVIOR*, defines the attitude of the customers towards returning used products and their response to company incentives aimed

Figure 3. Relationship between SERVICE AGREEMENT WITH CUSTOMER and return index for various CUSTOMER BEHAVIORS



at increasing the number of returns. The relationships among these three factors are shown in Figure 3. The difficulty in obtaining and documenting real data has led to the use of a distributional form constructed on intuitive grounds. However, a similar approach to the analysis of influences was adopted by Georgiadis and Vlachos (2004), who analyzed different parameters such as market behavior and the "green image" factor in relation to products in different industries.

In Figure 3 *return index* for a particular product is obtained based on the level of service that a company offers to its customers to encourage the return of the product after its use. In this regard, the values given to *SERVICE AGREEMENT WITH CUSTOMER* lie between 0 and 100%, which correspond to 0 and 1 of the *return index*, respectively. High values of service agreement are obtained for companies that offer incentives for the full return of sold products, such as through leasing contracts (e.g. cars and photocopiers) or service at the end of the useful life of a product/component (e.g. single-use cameras and toner cartridges). Also included are companies that maintain full responsibility for product recovery to abide by government environmental regulations. High values of service agreement correspond to high values of the *return index*, for which it is assumed that almost all sold products are returned by customers. The minimum value of service agreement corresponds to the kinds of products not involved in reverse logistics activity, particularly remanufacturing, resulting in a zero *return index* and no efforts by companies to generate product recovery. For all values in between, the relation between the index and service agreement depends on *CUSTOMER BEHAVIOR* (CB).

A range of company incentives are used in practice to stimulate a desired customer behavior in the area of product recovery (de Brito, Dekker, & Flapper, 2004), including: a deposit that must be paid when purchasing the product; free collection or repurchase of used products; a monetary incentive paid upon return of used products; and a trade-in that involves the possibility of obtaining a newer version of a product when the original product is returned. Currently, products designed for easier disassembly and clear information/advertising about reverse logistics activities and environmental responsibilities are being developed by companies to encourage the return of used products. However, the reaction to such incentives depends on individual customer behavior. To incorporate various possible customer behaviors into this study, three alternative relationships are assumed, which are represented in Figure 3. CB2 corresponds to a proportional relationship between *service agreement with customer* and *return index*. In this case, it is assumed that customers respond proportionally to incentives and services offered by companies attempting to recover used products. The symmetric curves CB3 and CB1 correspond to a quicker and a slower response from customers, respectively. Particularly in relation to CB3, it is assumed that the response of customers and consequently the associated *return index* changes quickly for low values of *service agreement with customer* while it is almost the same for higher values. This is different for CB1 which becomes more acute for higher values of *service agreement with customer*. This influence analysis is used to account for the relationship between customer behavior and quantity of returns, as much as is possible.

Mathematical Formulation

The dynamic relationships among the variables of the developed SFD for the production and inventory system for remanufacturing are defined by a set of mathematical equations. The symbology and the form used for the equations follow the conventions of the simulation software used to build the model: Vensim PLE v5.6d (Ventana Systems, Inc.). This provides a simple and flexible means of building simulation models from Stock and Flow Diagrams (Ventana Systems, Inc., 1999). The mathematical equations include several constant parameters or inputs. The latter correspond to the exogenous inputs of the SFD, which were represented in upper case letters, as well as to the initial values of the stock variables. Usually, when modeling a specific real system the values of these parameters are finetuned in order to reproduce the behavior of the system. However, because in this study we are developing a generic SD model of a production and inventory system for remanufacturing, assumptions are adopted for the value of the parameters. Indeed, given the objective of running a simulation of this generic model to evaluate strategies with a particular focus on the returns process, to improve the performance of the system we believe that an intuitive understanding regarding the impacts of the structure of the model on its dynamic behavior is more important than determining the exact value of the parameters. However, the assumptions used for the values of these parameters, which will be provided for the simulation analysis of the model, are set to correspond as far as possible to a meaningful concept based on real-world situations.

The mathematical meaning of an SFD results from the conventions used by Forrester (1961) to define this particular diagram based on the hydraulic metaphor of the flow of water into and out of reservoirs (Sterman, 2000). Specifically, stocks (reservoirs) accumulate material (water), which changes dynamically in response to the flow variables (inflow and outflow of water). For this reason, stocks can be represented mathematically through a differential equation (Sterman, 2000) or an equivalent integral equation. Thus, the stock variable *Collected Returns* is defined by a time integral of the net inflow (*collection*) minus the net outflows (*accepted returns* and *failed returns*):

$$\textit{Collected Returns}\ (t) = \int_{t_0}^{t} (\textit{collection}(t) - \textit{accepted returns}(t) - \textit{failed returns}(t))dt + \textit{Collected Returns}\ (t_0).$$

The *collection* flow is equal to the *returns* flow. This means that at time t, all returns follow a collection process: *collection (t) = returns (t)*. Infinite collection capacity is assumed as all possible returns are collected. *Failed returns* at time t are equal to total *Collected Returns* times the *PERCENTAGE DISPOSED* divided by the *INSPECTION TIME* (Vlachos et al., 2007). The percentage of disposed returns and the inspection time are assumed to be constant due to the difficulty of representing and modeling the real dynamic variance for this factor which depends on product characteristics, company quality policies and inspection strategies; and this particular issue is not within the scope of this study. *Accepted returns* at time *t* are the *Collected Returns* that pass the inspection process. For this reason, the percentage of returns accepted for remanufacturing is 1 – disposal percentage. Equations can thus be formulated as:

$$accepted\ returns(t) = \frac{\text{Collected Returns(t) * (1- PERCENTAGE DISPOSED)}}{\text{INSPECTION TIME}}$$

$$failed\ returns(t) = \frac{\text{Collected Returns(t) * PERCENTAGE DISPOSED}}{\text{INSPECTION TIME}}$$

An *IF THEN ELSE* function and the logical operator *AND* are used to define the production quantity in the system. In particular, they provide the number of production orders during the simulation period. The logical expression defines the condition when the *Serviceable Inventory* level is less than or equal to the *low level of serviceable for production* and also when the *Recoverable Inventory* level is less than the *remanufacture up to level* minus *low level of serviceable for remanufacturing*. If the condition is true, the expression returns a production order value equal to the ratio between *production up to level* minus the serviceable inventory on hand and the *replenishment frequency*; otherwise the returned value is zero. A similar equation defines the remanufacturing quantity and the number of remanufacturing orders in the model. In this case, the condition requires that the *Serviceable Inventory* level is less than or equal to *low level of serviceable for remanufacturing* and that *Recoverable Inventory* is greater than or equal to *remanufacture up to level* minus *low level of serviceable for remanufacturing*. The possible returned values are a remanufacturing order that is equal to the ratio between *remanufacture up to level* minus *low level of serviceable for remanufacturing* and *replenishment frequency* if the condition is true, or zero otherwise. *Recoverable* and *Serviceable Inventory* levels are defined in the equations below:

$$Recoverable\ Inventory(t) = \int_{t_0}^{t} (\text{acceptedreturns}(t) - remanufacturing(t))d + Recoverable\ Inventory(t_0)$$

The assumed values for the product demand are set in order to generalize the model for different kinds of products. Specifically, *demand* is set to a uniformly distributed random number. The random values are set at between 300 and 2,000 items with a fixed noise seed, in order to have the same sequence of random values for every simulation, which equals 2. The formulation of *demand* is thus represented as: *demand = RANDOM UNIFORM (300, 2,000, 2)*. *Demand inflow* represents the flow of previously sold products currently in use which are now used products and possible returns after the *residence time* has elapsed. In order to model this process the function *DELAY FIXED* is used. This function returns the value of the input *demand* delayed by the delay time, which in this case is the *residence time*. Zero is the initial value of *demand inflow* at the start of the delay process: *demand inflow = DELAY FIXED (demand, residence time, 0)*.

The flow of actual returned items that are collected is represented as a dynamic ratio between the proportion of *Used Products* through the use of *a returns rate* and the time required to return and collect the items:

$$returns(t) = \frac{\text{(Used Products(t) * returns rate(t))}}{\text{RETURN TIME}}$$

The variable *Used Products* is defined as:

$$Used\ Products(t) = \int_{t_0}^{t} (\text{demandinflow}(t) - returns(t))dt + Used\ Products(t_0)$$

Returns rate represents the proportion or percentage of used products that are returned during the time period under consideration. Several authors, such as Kiesmuller (2003), Kiesmuller and Minner (2003) and Inderfurth (2005), use the returns rate variable in their models. In order to define the quantity of returns, they consider a returns rate to be the ratio between the average returns and the average demands. Consequently, the *returns rate* in this model is represented as a dynamic ratio between *returns inflow* and *demand*:

$$\text{Returns rate (t)}\ \frac{\text{returns inflow(t)}}{\text{demand(t)}}$$

The function *DELAY FIXED* is used to identify the value of the input given by the previous ratio delayed by the *residence time* plus one time period. The reason for this delay is due to the necessary time equivalence between the variables *returns* and *returns rate*, as the accumulation of used products and the actual returns flow start one time period after the residence time. *Returns inflow* represents the expected returns of demand. A forecast of returns is obtained using the *return index*: *returns inflow (t) = demand (t) * return index (t)*.

A functional relationship between two variables is used for the formulation of the *return index* at time *t*. This is obtained by using a lookup function which allows the definition of a customized relationship between a variable and its causes. Specifically, an equation gives the value of *return index* at any value of *SERVICE AGREEMENT WITH CUSTOMER* through a linear interpolation between the values specified in *return index lookup* (Figure 2) as: *return index = return index lookup (SERVICE AGREEMENT WITH CUSTOMER)*. *Return index lookup* defines the lookup function that expresses the relationship between *return index* (dependent variable) and *SERVICE AGREEMENT WITH CUSTOMER* (independent variable), as shown in Figure 3. For this reason, the variable *return index* is formulated through a combination of *IF THEN ELSE* and lookup functions: *return index (t) = IF THEN ELSE (CUSTOMER BEHAVIOR = 3, return index lookup 3 (SERVICE AGREEMENT WITH CUSTOMER), IF THEN ELSE (CUSTOMER BEHAVIOR = 2, return index lookup 2 (SERVICE AGREEMENT WITH CUSTOMER), return index lookup 1 (SERVICE AGREEMENT WITH CUSTOMER)))*. This equation represents the tendency of a particular product to be returned by customers, considering individual customer behaviors and differing levels of service agreement or company incentives. The constant *CUSTOMER BEHAVIOR* can assume three values, 1, 2 or 3, represented by the three different curves in Figure 3. In the function of value assumed for *CUSTOMER BEHAVIOR*, *return index* is calculated through one of the lookup functions.

Finally, the accumulation of returns from *returns inflow* is defined through the time integral of the flow:

$$Returns\,Accumulation(t) = \int_{t_0}^{t} (\text{returnsinflow(t)})\,\text{dt} + Returns\ Accumulation\ (t_0)$$

Validation Analysis

Before we can run a simulation of the model for evaluating and investigating strategies to improve the performance of the system, a validation analysis must be performed to determine whether the model is suitable for this objective. In SD theory, model validation primarily involves the assessment of the structure and behavior of the model in terms of its consistency with the available facts and descriptive knowledge of a real-world system (Morecroft, 2007; Sterman, 2000). Such assessment, as Sterman states, is useful to build confidence that a model is appropriate for its purpose.

Although the theoretical basis of the quantitative modeling in this study was obtained from the existing literature as well as from some of the information and data collected from the companies employed as case studies, in order to develop a model that corresponds to a meaningful concept representing the real world (Sterman, 2000), we deemed it necessary to perform validation tests to define its capacity to reflect the structure and behavior of a real process model.

The validation tests were the *direct structure tests* and *structure-oriented behavior tests* (Barlas, 1996). Specifically, for the *direct structure tests*, in which simulation was not involved, we engaged extreme condition tests in order to check whether each mathematical equation of the model made sense and was reasonable given the available knowledge of the real system. For the *structure-oriented behavior tests*, which involved simulation of the entire model, we undertook behavior sensitivity tests. These involved sensitivity analysis on particular parameters of the model aimed at comparing the high sensitivity of these parameters between the model and a real system.

Based on the outcome of the validation analysis we determined that the developed model of the production and inventory system for remanufacturing is suitable for evaluating and investigating strategies aimed at improving the performance of the system through a simulation of scenarios. Specifically, through the direct structure validation we confirmed the capacity of the model equations to be logically and dimensionally consistent with the available knowledge of real-world scenarios. Through the sensitivity analysis, we also identified that the changes in the behavior patterns of the model, due to particular assumptions regarding the changes in value for particular parameters, are consistent with a real system.

Another outcome of the sensitivity analysis was the identification of those parameters to which the model showed sensitivity. In particular, in line with the purpose of the model, it was found that the model showed sensitivity to the changes or joint changes in the values of the main parameters involved in the returns process (*RESIDENCE TIME*, *SERVICE AGREEMENT WITH CUSTOMER* and *CUSTOMER BEHAVIOR)*. Thus, these parameters can be used for the simulation analysis of scenarios to achieve the aims of the model.

SIMULATION OF SCENARIOS

The development of the model, and its validation, led to the final stage of SD simulation modeling, which involved simulation of scenarios focusing on the main parameters of the returns process in order to reach conclusions, specifically to identify and evaluate the best policy and strategy to adopt, and what occurs in the system if factors change or events intervene. In particular, the main purpose of the simulation is not to create predictions or forecasts of a future event, but rather to evaluate scenarios or alternative futures that could occur given certain assumptions or conditions (Morecroft, 2007).

Before the various scenarios can be designed and simulated, we must first identify and present the measure of performance and the base scenario used for the simulation analysis. We then discuss the various scenarios employed in the analysis and present the results of the simulations from which an evaluation can be undertaken of the best strategies to adopt to improve system performance.

Performance Measure

The measure of performance for this simulation analysis represents the output variable the value changes of which under different scenarios and given certain values of the parameters enhance our understanding of the conditions that might arise in the system. These conditions which represent the simulation results lead to an evaluation of strategies, which in turn provide guidelines as to how to improve the performance of the system.

As was found in the existing literature, for most production and inventory systems for remanufacturing the objective is to minimize the total inventory cost; therefore, the latter will be considered as a measure of performance for the simulation analysis. However, the analysis is not aimed at determining the optimal order quantity or reorder inventory level. Rather, the objective is to identify the effects on the system of the main factors involved in the returns process, in particular residence time, service agreements offered by companies and customer behavior. Changes in these factors are considered to be events that intervene in a production and inventory system for remanufacturing where the returns process is characterized by uncertainty around the quantity and timing of returns.

The value of the total inventory cost for the analysis is obtained by adding a number of operational costs. These are the set-up costs for each production order and remanufacturing order, the cost of stockout for each out-of-stock sale, and the holding costs for recoverable and serviceable inventory. The choice of this particular sum of total cost in which the remanufacturing and production costs are excluded is based on several observations. First, one of the assumptions previously mentioned in the description of the system under study determines that the model posits a disposal activity resulting from the inspection process rather than a planned disposal in recoverable inventory. This means that all accepted returns from the inspection are stored as recoverable inventory and used for the remanufacturing activity. Therefore, there are no inventory decisions that affect the remanufacturing and production activity in terms of quantity of items to be remanufactured and produced (van der Laan & Teunter, 2006). Moreover, the mathematical formulation of the model considers the number of remanufacturing orders and production orders rather than production and remanufacturing rates. These observations render the system independent of the remanufacturing and production activity, and consequently it excludes the production and remanufacturing costs. However, in order to potentially consider cheaper remanufacturing in place of production activity (remembering that the cost of remanufacturing is typically 40–60% of the cost of production), different set-up costs are assumed for production orders and remanufacturing orders. Moreover, this analysis focuses on the total inventory costs which exclude the collection and inspection/disposal costs.

Base Scenario

The base scenario represents and determines the values of the parameters that we initially set in order to run the simulation of scenarios. Specifically, these scenarios are obtained by changing the value of the main parameters under study while at the same time keeping the other parameters at the value of the base scenario.

The base scenario used for this simulation analysis involved the use of several assumptions aimed at ensuring the model corresponds as much as possible to a meaningful representation of the real world. Such assumptions reflect the theoretical basis drawn from the existing literature and to a degree information obtained from the companies employed as case studies. Indeed, the generic models based on remanufacturing and closed loop supply chains presented in the literature use several assumptions in relation to the values of the model parameters. For this reason, we deemed it appropriate to adopt similar assumptions for the purposes of the simulation analysis.

In setting the values of the model parameters, *INSPECTION TIME*, *REPLENISHMENT FREQUENCY* and *RETURN TIME* were set to one month. It was assumed that all returns collected in a given month are inspected within that same month, that the remanufacturing and production orders are replenished monthly, and that there is a monthly collection of returns. The initial values at the beginning of the simulation horizon for the stock variables were set to zero (Georgiadis & Vlachos, 2004) for *Collected Returns*, *Recoverable Inventory* and *Used Products*, while 2,000 items were set for the *Serviceable Inventory*. *PERCENTAGE DISPOSED* was set at 0.1 (10%), 500 items for *low level of serviceable for remanufacturing* (*sr*), 4,000 items for *REMANUFACTURE UP TO LEVEL* (*Sr*), 4,000 items for *PRODUCTION UP TO LEVEL* (*Qm*), 300 items for *low level of serviceable for production* (*sm*) and 12 months for the *RESIDENCE TIME*. *CUSTOMER BEHAVIOR* was set at 2, with a proportional relationship between *return index* and *SERVICE AGREEMENT WITH CUSTOMER*, which in turn assumed a value equal to 50%. In defining the values of the parameters related to the measure of performance, set-up costs were set at $20 per remanufacturing order and $50 per production order. The cost of stockout was set at $10 per out-of-stock sale, with $0.5 and $0.8 per unit recoverable and serviceable inventory holding cost, respectively. However, the generality of the model provides the opportunity to finetune and customize the values chosen for the basic scenario for different kinds of products and different industries.

Scenarios Derived from the Returns Process

The simulation of scenarios focused on the main parameters considered for the returns process. Thus, the scenarios were obtained through the combination of a range of values for those parameters related to the system policies defined for the returns process. Moreover, the model was sensitive to these parameters during the sensitivity analysis. Specifically, these parameters are *RESIDENCE TIME* (RT), *SERVICE AGREEMENT WITH CUSTOMER* (SAWC) and *CUSTOMER BEHAVIOR* (CB).

The analysis focused on the effect of the three parameters on the total inventory cost during the planning horizon, which was set at 60 months. Specifically, through this measure of performance changes in the behavior of the modeled production and inventory system for remanufacturing were examined using 6 levels of *residence time*, 5 levels of *SERVICE AGREEMENT WITH CUSTOMER* and 3 levels of *CUSTOMER BEHAVIOR*. Figure 4 lists the values of the parameters used for the analysis, which involved a total of 90 scenarios. The selection of the assumed values for the parameters was intended to correspond as much as possible to a meaningful reflection of the real world for a broad range of products. However, following the purpose of this simulation analysis to evaluate strategies aimed at improving the performance of the system, we believe that the exact value of the parameters is not as important as an understanding of the changes in the behavior of the system under different scenarios.

The various scenarios are characterized by low and high *residence times*, which correspond with *fast-used products* and *slow-used products*, respectively. The choice of this table structure was based on the relationship between *residence time* and type of product. *service agreement with customer* is considered to be the policies/incentives that companies use to retrieve used products and *customer behavior* the customer tendency to return them, respectively. The parameter values used to set the residence time are realistic as they can be associated with several remanufacturable products (Georgiadis et al., 2006). The assumed values for *service agreement with customer* represent a broad range of company policies and incentive types that develop a relationship between companies and their customers in the returns process. In the same way, the different levels of *customer behavior* are representative of a broad range of responsive aptitudes.

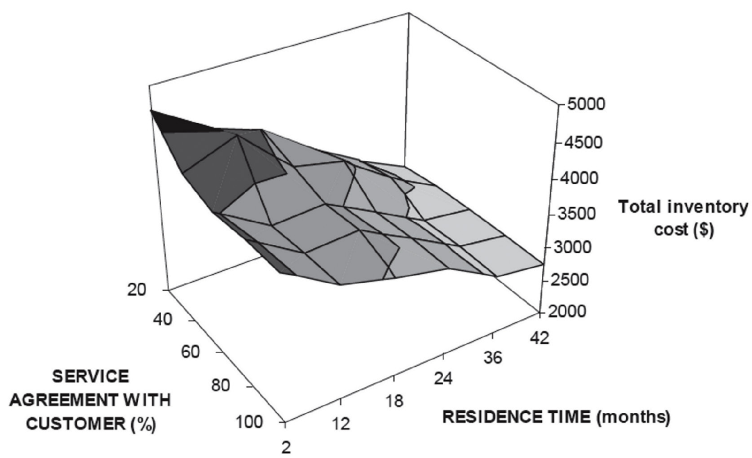
Figure 5 represents the evolution of the total inventory cost for various simultaneously simulated levels of *residence time* and *service agreement with customer*, given a *customer behavior* level equal to 3.

The first observation shows a decrease in the total inventory cost for a high level of *residence*

Figure 4. Parameter values used for returns process scenarios

Scenarios	RESIDENCE TIME (months)	SERVICE AGREEMENT WITH CUSTOMER	CUSTOMER BEHAVIOR
Base scenario	12	50%	2
Fast-used products	2	20%, 40%, 60%, 80%, 100%	1, 2, 3
	12	20%, 40%, 60%, 80%, 100%	1, 2, 3
	18	20%, 40%, 60%, 80%, 100%	1, 2, 3
Slow-used products	24	20%, 40%, 60%, 80%, 100%	1, 2, 3
	36	20%, 40%, 60%, 80%, 100%	1, 2, 3
	42	20%, 40%, 60%, 80%, 100%	1, 2, 3

Figure 5. Evolution of total inventory cost changing RT and SAWC



time. This is due to the reduced cyclic nature of returns for *slow-used products*, which remain with customers for longer. In the model the reduced cyclic nature of the return of a product is represented by the long time period between the product sale and its possible return. Specifically, a portion of products sold at time 0 of the planning horizon become remanufacturable returns only after a long residence time, and during this residence time they do not affect the recoverable inventory and its associated cost. Conversely, products with a short residence time engender a fast cyclic nature of return and are quickly involved in recoverable inventory and remanufacturing activity. Therefore, in the short time period, s*low-used products* lead to a reduced use of recoverable inventory and remanufacturing activity with a consequent reduction in the total inventory cost. This reduction is also due to the serviceable holding costs, which, according to our analysis, resulted indifferent to the remanufacturing or production of the product and consequently of the residence time, and stockout costs were often special events.

Our aim is not to reveal that *slow-used products* are promising candidates for profitable remanufacturing systems. On the contrary, in the context of a closed loop supply chain, examples of profitable remanufacturing processes include *fast-used products* such as single-use cameras and assemblies or sub-assemblies of copiers/printers. These kinds of products generate high levels of return and recoverable inventory with a subsequent increase in the total inventory costs. At the same time, cheaper remanufacturing activity can then be adopted as a substitute for more expensive production activity, as was shown by the reduction of the total set-up cost during the simulation analysis. As the structure of the model does not involve remanufacturing and production costs then it is not possible to draw any conclusions regarding the profitability of remanufacturing. However, several examples have been provided in the literature of products with low residence times or short lifecycles and high returns rates for which high stock levels increase inventory costs. At the same time, in the long run these products generate a more profitable remanufacturing activity due to reduced production costs such as purchasing and service costs (Flapper, Van Nunen, & Van Wassenhove, 2005).

The second observation obtained through the scenarios simulation reinforces the possibility that a high returns rate can generate a more profitable remanufacturability in closed loop supply chains. From Figure 5 it is possible to observe a decrease in the total inventory cost for each level of *residence time* except for the highest one, by increasing the *service agreement with customer* from 20% to 100%. The cost variation is not as significant as that which occurs when increasing the *residence time*. However, it can improve efficiency in managing inventory in the remanufacturing process. High incentives for product recovery and consequently a high returns rate and quantity of remanufacturable returns can increase the level of recoverable inventory, which can be used to generate remanufacturing orders faster as a substitute for production. This reduces the average level of recoverable inventory and consequently the total inventory cost, as was shown by the reduction of the recoverable holding costs during the simulation analysis. Moreover, increasing remanufacturing activity does not negatively impact the effectiveness of the system, as stockout quantity and costs showed a reduction for a higher returns rate.

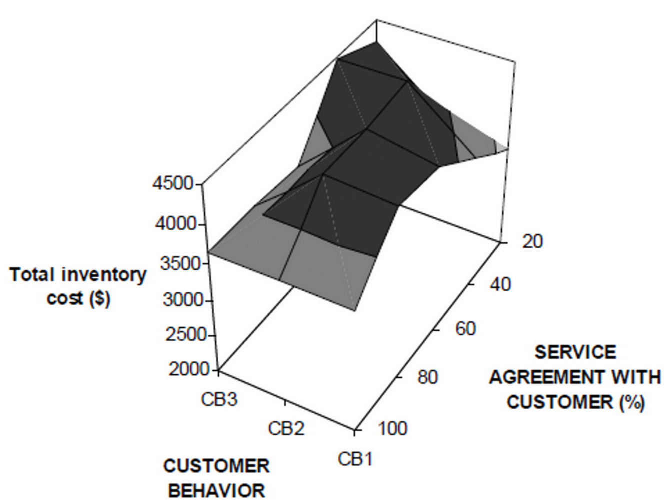
The simulation analysis showed a higher reduction of the total inventory cost, increasing the *service agreement with customer* from 20% to 100%, for *fast-used products* than for *slow-used products* for which an increase in cost characterizes the highest residence time (42 months). This difference in cost trend is due to the lesser influence of *slow-used products* on recoverable holding costs and recoverable inventory, as noticed in the first observation. In this case, an increase in the service agreement or incentives increases the quantity of recoverable inventory but only after a long residence time. This surplus of recoverable inventory does not affect remanufacturing as a substitute for production activity in the short term. Therefore, an increase in service agreement for product recovery could have a reduced or negative effect on total inventory cost for *slow-used products* over a short time period. This is different for *fast-used products*, where an increase in incentives affects the quantity of recoverable inventory in a shorter time period, which can then be used sooner in remanufacturing activity with subsequent benefits in recoverable inventory and production activity reduction. However, in the long term the return of *slow-used products* can be improved by a percentage increase in the service agreement. Increasing the planning horizon from 60 to 120 months, the total inventory cost for *slow-used products* with a *residence time* equal to 42 months decreases from $3,834 for 20% of *service agreement with customer* to $3,601 for 100% of *service agreement with customer*. Examples of closed loop supply chains for *slow-used products* such as white goods are presented in the literature (Flapper et al., 2005). The main driver of the reverse logistics process for such products is government legislation, which stipulates the responsibilities of producers to recover their end-of-life products. However, incentives to the customer and several other factors involved in the process are given in order to increase the returns rate and improve economic and environmental benefits.

Simulations, using the same values for the parameters *RESIDENCE TIME* and *SERVICE AGREEMENT WITH CUSTOMER* presented in Figure 4, were undertaken by setting *CUSTOMER BEHAVIOR* at equal to 2 and 1. Figure 6 shows the evolution of the total inventory cost for changes in *customer behavior* and percentage of *service agreement with customer,* with a *residence time* equal to the base value of 12 months.

A reduction in the value of *customer behavior* from 3 to 1 decreases the returns rate and consequently the average number of returns. This applies to every percentage value of *service agreement with customer* except for 100%, for which the *return index* and consequently the returns rate is found to be independent of customer behavior. These simulation results correspond with and are thus representative of real-world scenarios, as a lower response from customers to company incentives for the recovery of used products in a closed loop supply chain system can reduce the quantity of returns. However, this does not tend to occur for products for which companies maintain ownership such as products under leasing contracts, or have recovery responsibilities as determined by government legislation.

This reduction in returns quantity resulting from changes in customer behavior has several consequences on the system and its total inventory cost. From Figure 6 it is possible to observe that for higher percentages of *service agreement with customer*, such as 60% and 80%, the total inventory cost increases if the customer response (and returns rate) is lower. This is due to the lower level of recoverable returns in the short term. Higher levels of service agreement or incentives coupled with higher response levels from customers to these incentives increase the level of remanufacturable returns. Therefore, in a shorter timeframe it is possible to use remanufacturing as a substitute for production activity with subsequent economic benefits obtained from the reduction of recoverable inventory and holding costs and the cheaper manufacturing processes. This observation was confirmed by lower recoverable holding costs obtained in a shorter period

Figure 6. Evolution of total inventory cost changing SAWC and CB



within the simulation analysis. Specifically, it was observed that when the remanufacturable returns start to accumulate, the recoverable holding costs are higher for the higher value of *customer behavior* due to the greater quantity of recoverable inventory. However, the latter allows for a prompt remanufacturing activity, which in turn reduces the recoverable inventory level and the costs. The same observation has been previously noted for an increase in *service agreement with customer*.

From Figure 6 it is evident that for lower percentages of *service agreement with customer*, such as 20% and 40%, the total inventory cost decreases if the *customer behavior* value decreases. In this case, as the analysis revealed, the low quantity of returns due to low incentives and low customer response leads to a low level of recoverable inventory, which is almost always lower than the level of recoverable inventory for higher values of *customer behavior*. The latter involves more remanufacturing activity but not enough to reduce the recoverable inventory level. Therefore, in the case of low values for *customer behavior* the level of recoverable inventory involves lower holding costs but at the same time lower remanufacturing activity. This scenario could negatively impact companies involved in remanufacturing activity for closed loop supply chain systems which require a sufficient quantity of returns to increase remanufacturing as a substitute for production activity. However, the analysis showed that in the case of low *customer behavior* the increased percentage in the service agreement might not generate an expected reduction in cost. As already mentioned, this is due to the insufficient level of remanufacturing activity needed to reduce and lower recoverable holding costs.

Evaluation and Recommendations

Remanufacturing activity in closed loop supply chain systems requires an adequate quantity of remanufacturable returns in order to establish a manufacturing process in which cheaper remanufacturing can be used as a substitute for production activity. This process leads to economic and environmental benefits in reducing the reliance on the more costly production activity. Remanufacturing uses 85% less energy than production, and reduces landfill, pollution and raw material usage (Gray & Charter, 2007). Moreover, the simulation analysis revealed that an increase in remanufacturing activity can optimize the inventory system and its costs through enhanced efficiency in the management of recoverable inventory. *Slow-used products* with a longer residence time in the short term present a reduced use of recoverable inventory due to their slower cyclic nature of return and consequently lower inventory cost compared to *fast-used products*. However, in the short run this could negatively affect remanufacturing as a substitute for production activity due to a shortage of remanufacturable returns. On the other hand, *fast-used products* can be used within a shorter time period and therefore prompt remanufacturing activity, which reduces the inventory cost through greater efficiency in recoverable inventory management. A prompt remanufacturing activity depends on the recoverable inventory on hand which, as the simulation analysis showed, is also influenced by *SERVICE AGREEMENT WITH CUSTOMER* and *CUSTOMER BEHAVIOR*. An increase in both these parameters leads to a higher returns rate, which in turn generates a higher level of recoverable inventory on hand and increases the possibility of prompt remanufacturing activity. For these reasons, the evaluation suggests that, for companies involved in remanufacturing activity as a substitute for the more expensive production activity, a shorter residence time combined with an increased level of service agreement or incentives for product recovery and a higher response from customers to these incentives can generate economic benefits through an increase in the quantity of remanufacturable returns.

However, uncertainty around the returns flow, particularly related to the timing and quantity of returns, could influence the results of the previous analysis of the inventory system. Some companies manage this uncertainty around the quantity of remanufacturable returns stored in the recoverable inventory without attempting to balance returns with demands, preferring instead to dispose of excess inventories on a periodic basis (Guide, 2000). In the same way, several authors incorporate a planned disposal of recoverable inventory in their models and posit a simple probabilistic returns quantity or define all demands as returns (van der Laan, Dekker, & Salomon, 1996; van der Laan & Salomon, 1997; Vlachos et al., 2007). In our model, planned disposal of recoverable inventory is not considered, which could add a new inventory cost and be more profitable, following the reverse logistics *preponement* concept (Blackburn, Guide, Souza, & Wassenhove, 2004), during or prior to the inspection stage. However, uncertainty around the timing and quantity of returns is specifically tackled in this study through the use of parameters such as *Residence time*, *SERVICE AGREEMENT WITH CUSTOMER* and *CUSTOMER BEHAVIOR*. Analysis of their interrelationships in the returns process can provide a forecasted returns rate and a possible time of return for used products with different product characteristics and in different industries. Knowledge of a product's residence time coupled with combinations of incentives in product recovery such as trade-in and leasing contracts can also assist in estimating the time and quantity of returns. Moreover, incentives, particularly leasing contracts and changes in product design for easier disassembly and recovery of product/components, can result in a reduction of the residence time, and subsequent benefits as those previously mentioned in connection to *fast-used products*. For example, the introduction of leasing contracts, or changes in product design that enable customers to easily disassemble and return used products/components, can fix or reduce the residence time. Through such incentives, companies can influence customer behavior in returning used products. By adopting policies such as the introduction of deposit fees, free collection or repurchase of used products, fees paid upon the return of used products, and improving product design, alongside providing clear information/advertising about reverse logistics activities and environmental responsibilities, can assist in enhancing customer behavior in relation to the returns process.

CASE STUDIES

In this section we assess the robustness of the findings obtained through the simulation analysis, which leads to the evaluation of the strategies aimed at improving the performance of the system. For this purpose, we use the data and information collected from the two companies employed as case studies: the Australian Mobile Telecommunications Association (AMTA) and Fuji Xerox Australia. These companies are involved respectively in reverse logistics and remanufacturing activity, and in particular are engaged in returns process activities, which meant they were appropriate cases for our study.

The findings were obtained from a generic model of the production and inventory system for remanufacturing. For this reason, the assessment involves a comparison between the findings and the data and information collected from these companies in order to find similarities and to verify the former by referring to the actual real-world practices of these companies. Specifically, the quantitative data and information collected from AMTA are employed first. Then the similarities between the AMTA data and both the qualitative data and the information collected from Fuji Xerox Australia are identified and discussed.

Australian Mobile Telecommunications Association (AMTA)

The first Australian company case is examined to assess the findings that revealed that an increase in the service agreement or incentives for product recovery, which in turn increases the likelihood of customers to engage in the returns process, leads to a higher returns rate.

AMTA is the national body of the mobile telecommunications industry in Australia and is currently supporting the official national recycling program of the mobile phone industry (KPMG, 2008). This program is a free recycling project for mobile phone users, which avoids landfill activity and recovers material from used mobile phones, bringing environmental and economical benefits. Project managers were contacted in order to obtain data and information, not about the recycling program (which is not directly relevant to this study) but about the influence of incentives on customers to return used products, and on customer behavior and returns rates in general. Data and information were obtained from the organization's 2007–2008 annual report (KPMG, 2008).

According to the report, the company has developed numerous incentives and service agreements with customers, retailers and other reverse logistics actors such as local councils and recyclers in order to increase returns/collection rates and improve customer behavior in relation to the returns process. In particular, the focus of the program is on free used product collection from customers. This has been achieved by distributing reply paid recycling satchels available in selected mobile phone packs and by setting up public collection points nationwide in retailers and Australia Post outlets. Other incentives include customer communications and environmental campaigns about the program published in catalogues, on websites, through direct marketing and television advertising presented by mobile manufacturers, service centers and retailers. These activities have achieved varying results. In particular, it was interesting to notice the evolution of particular key performance indicators for the program for the years 2005 to 2008. Since 2005, awareness of the recycling program, which represents consumer attitudes towards the reverse logistics program, has increased from 46% to 75% and at the same time collection and collection rates have increased from 42 tons to 97 tons and from 15% to 19%, respectively. The latter two factors are representative of the quantity of returns and the returns rate of used mobile phones. Moreover, the disposal to landfill rate has decreased from 9% to 4%.

Therefore, similar to our findings, these company incentives have enhanced customer behavior in relation to the returns process, which in turn has led to an increased returns rate.

Fuji Xerox Australia

The case study of Fuji Xerox Australia is presented in this section in order to assess two of the findings obtained from the simulation analysis. The first regards our belief that benefits can be obtained from remanufacturing activity through the combination of a shorter residence time and an increased level of service agreement or incentives for product recovery. The second regards the significant influence of service agreement or incentives for product recovery on the uncertainty around the quantity and timing of returns.

This case study employs qualitative data and information on this world leader in remanufacturing processes, obtained from the company management and drawn from the existing literature. Specifically, the qualitative data and information refer mainly to the Eco Manufacturing Centre located in Sydney.

The Eco Manufacturing Centre is the distribution center for remanufactured printer and copier assemblies or sub-assemblies, which would otherwise be landfill for the Asia-Pacific Region (Fuji Xerox Australia, 2007). Assemblies and sub-assemblies, removed from equipment during maintenance service calls, are remanufactured at this center. According to the managers, the Eco Manufacturing Centre focuses on and deals only with remanufacturing activity. For this reason, returns are an essential element in its remanufacturing activity, and predicting the quantity and timing of returns is essential to ensuring an efficient inventory and planning/scheduling for remanufacturing.

The information obtained from the managers is used to assess the first research finding. Specifically, the introduction of changes in product design as an incentive for product recovery through the easier disassembly and recovery of components rather than the whole machine, which in turn leads to faster recovery of returns or shorter residence times, has been strongly supported by the Fuji Xerox managers. Indeed, new machines are built for easier and faster disassembly, recovery and remanufacturing processes. In this way, remanufacturing activity can focus on modules that have shorter residence times and easier remanufacturability. The company's aim is to increase the number of remanufacturable returns and maximize engagement in the more profitable remanufacturing activity as a substitute for production activity. The work of the Eco Manufacturing Centre, which deals only with remanufacturing activity, supports this aim.

In terms of the second research finding, Fuji Xerox managers believe that a *full service agreement with customer* is an important strategy in the remanufacturing process. Fuji Xerox draws up a full service and maintenance agreement with its customers, whose preference is usually to lease the products. The full service agreement has a number of targets, including a marketing strategy aimed at increasing service levels for the customer. However, from the remanufacturing activity point of view the full service agreement is also a returns process strategy which functions to increase control of the quantity and timing of returns and to improve remanufacturing activity. The service is conducted by engineering teams, who provide service and repairs for breakdowns or when customers find the product is not working satisfactorily, as well as preventive maintenance of products. Using diagnostic tools, this maintenance service identifies the reasons for failure and opportunities to extend the product life (Fuji Xerox, 2007). This process involves two main analyses:

- **Failure Mode Analysis:** to identify the reasons for the failure of assemblies and sub-assemblies.
- **Signature Analysis:** to determine the remaining life of the assemblies and sub-assemblies through an examination of their critical performance parameters.

Data collection on these processes has led to continuous improvement of the basic product design and has resulted in a number of improvements. The durability of main assemblies and sub-assemblies, which is usually measured on the basis of the possible maximum number of copies to be completed, is designed so as to enable replacement of all of the components at the same time. In this way, during preventive maintenance it is possible to predict when it will be necessary to replace the assemblies to improve product performance, and when it will be necessary to replace them altogether. This process is guided by computer systems that identify whether or not there is a remanufacturing program for an assembly. During the breakdown and preventive maintenance, assemblies and sub-assemblies are replaced, and information on problems, solutions and forecasts for future replacement is provided. If the replacement is not possible at the maintenance stage because there is no ready availability of new assemblies, an order is submitted to the local warehouse. Used assemblies and sub-assemblies are collected by the service engineers, valet service staff or dealers to be returned and stored at the local warehouse. From the local warehouses, they are transported to the Eco Manufacturing Centre to be re-engineered and remanufactured as products that are good as new. Finally, they are packaged and stored in the surrounding warehouse as new products, awaiting transportation to the central distribution center of Fuji Xerox Australia.

From the information collected we can conclude that one of the reasons why the Fuji Xerox managers at the Eco Manufacturing Centre strongly support the introduction of a full service and maintenance agreement with customers is that it will increase control of the quantity and timing of returns. Indeed, the use of leasing contracts for products, combined with a service agreement that can generate a forecast regarding the quantity and timing of returns, allows the company to keep track of the quantity and location of equipment and maintain a degree of control over the returns rate.

FUTURE RESEARCH DIRECTIONS

The study described in this chapter explored a series of issues relevant to the field of reverse supply chains and in particular closed loop supply chains. However, we believe there is scope for further lines of enquiry in this area, to be considered by future researchers.

The findings of this study were obtained through the modeling and simulation of a generic production and inventory system for remanufacturing which is generally in line with those used in previous studies. In order to simplify the analysis and interpretation of such a system, a number of assumptions were made. However, an opportunity for further research lies in the evaluation of system performance by relaxing the remaining assumptions. This could be achieved by remodeling the structure of the system and incorporating the factors and their influence relationships that affect the system activities/processes.

Improving the model structure and its applicability to the real world might also be achieved through an extension of the meaning of the existing factors. Specifically, several exogenous factors such as service agreement with customers, customer behavior and residence time were defined in a generic way without considering their social, economic or management backgrounds. An exploration of the origins and roles of these factors could expand the models to cover new areas of research focusing on different fields of study. For example, in the returns process the factors residence time and service agreement with customers could be connected with perspectives from the social, legislative and marketing disciplines. Therefore, this study could be the starting point for further research aimed at analyzing the impact that such perspectives and considerations may have on a reverse supply chain system.

CONCLUSION

In this research, we developed an SD simulation model of the production and inventory system for remanufacturing within the context of closed loop supply chains. Our focus, particularly for the simulation analysis, was on the returns process within the system in which the returns rate was modeled by manipulating the relationships between particular factors that affect the system. The selected factors identify the time period for which products stay with customers, or residence time, and the quantity of possible returns based on customer demand or the return index. In particular, the return index was obtained by considering the relationship between the company incentives or service agreement with the customer aimed at encouraging used products recovery, and the customer behavior in terms of product return. The remanufacturing process was also modeled and a pull inventory control policy was applied to the process.

By analyzing the total inventory cost as a measure of the performance of the system, several findings were obtained regarding the effects of residence time and changes in the level of company incentives and the resulting customer behavior. The main finding is that companies engaged in remanufacturing activity can enhance their efficiency in managing inventory through shorter residence times and an increased level of company incentives, which results in improved customer behavior. This leads to a higher level of recoverable inventory on hand and consequently the possibility of prompt remanufacturing as a substitute for the more costly production activity, which in turn reduces the recoverable inventory level and its related holding cost.

Moreover, company incentives for the recovery of used products have significant influences on the uncertainty around the quantity and timing of returns and ultimately on total inventory costs. Increasing company incentives or enhancing service agreements with customers, which in turn improves customer behavior in returning used products, can improve the control of returns. Two company case studies (the Australian Mobile Telecommunications Association and Fuji Xerox Australia) were employed to assess the findings. It was found that Fuji Xerox Australia in particular offers incentives such as changes in product characteristics, full service and maintenance agreements with customers, and leasing contracts that function to reduce the product residence time and improve the company's control over returns.

Following our purpose to model a production and inventory system for remanufacturing within the context of closed loop supply chains and to evaluate effective control strategies aimed at improving system performance, this chapter made several contributions in different research areas. Specifically, the modeled returns process can be linked to specific product categories using the knowledge of the distinctive elements of particular products, which in this case was the average residence time. The latter characterizes different types of products depending on their variable time of use and recovery time, which in turn affects the system through the timing of returns. Service agreements with customers can influence customer behavior in relation to returning used products, and consequently affect the quantity of returns within the system. The variability of customer behavior, which generates uncertainty around the quantity of returns within the system, has been modeled and analyzed through three different customer behavior patterns. Our contribution in this area is to extend the current research with an approach to product recovery in which the correlation between demand and returns is identified through the use of these particular factors.

On the basis of the findings obtained from the simulation analysis, our contribution is to offer observations regarding efficiency in managing inventory activities for this particular system, and to provide guidelines for determining the quantity and timing of the return of used products by customers in order to reduce the uncertainty surrounding the timing and quantity of returns.

REFERENCES

Angerhofer, B. J., & Angelides, M. C. (2000, December). *System dynamics modelling in supply chain management: research review.* Paper presented at the 2000 Winter Simulation Conference, Orlando, FL.

Automotive Parts Rebuilders Association. (2007). *The remanufacturing industry in Europe*. Retrieved May 15, 2007, from http://www.apra-europe.org/ main.php?target=industryinformation

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, *12*(3), 183–210. doi:10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4

Biehl, M., Prater, E., & Realff, M. J. (2007). Assessing performance and uncertainty in developing carpet reverse logistics systems. *Computers & Operations Research*, *34*(2), 443–463. doi:10.1016/j.cor.2005.03.008

Blackburn, J. D., Guide, V. D. R. J., Souza, G. C., & Wassenhove, L. N. V. (2004). Reverse supply chains for commercial returns. *California Management Review*, *46*(2), 6–22.

Blanc, H. M. I., Fleuren, H. A., & Krikke, H. R. (2004). Redesign of a recycling system for LPG-tanks. *OR-Spektrum*, *26*(2), 283–304. doi:10.1007/s00291-003-0145-3

Blumberg, D. F. (2005). *Introduction to management of reverse logistics and closed loop supply chain processes*. Boca Raton, FL: CRC Press.

Coyle, R. G. (1996). *System dynamics modelling: a practical approach*. London: Chapman & Hall.

de Brito, M. P., & Dekker, R. (2003). Modelling product returns in inventory control: exploring the validity of general assumptions. *International Journal of Production Economics*, *81/82*, 225–241. doi:10.1016/S0925-5273(02)00275-X

de Brito, M. P., Dekker, R., & Flapper, S. D. P. (2004). *Reverse logistics: a review of case studies*. In T. P. R. European Commission (Ed.), European Commission, TMR project REVELOG.

Dekker, R., Fleischmann, M., Inderfurth, K., & Wassenhove, L. N. V. (2004). *Reverse logistics: quantitative models for closed-loop supply chains*. Berlin: Springer.

Environment Protection Authority. (2002). *Case study: making products better than new*. Sydney, NSW: Environment Protection Authority.

Eoksu, S., Sungwon, J., Haejoong, K., & Jinwoo, P. (2004). *Genetic and evolutionary computation*. Heidelberg: Springer.

Flapper, S. D. P., Van Nunen, J., & Van Wassenhove, L. (2005). *Managing closed-loop supply chains*. Heidelberg: Springer. doi:10.1007/b138818

Fleischmann, M., Bloemhof-Ruwaard, J. M., Dekker, R., Van der Laan, E., Nunen, J. V., & Wassenhove, L. V. (1997). Quantitative models for reverse logistics: a review. *European Journal of Operational Research*, *103*(1), 1–17. doi:10.1016/S0377-2217(97)00230-0

Forrester, J. W. (1958). Industrial dynamics: a major breakthrough for decision makers. *Harvard Business Review*, *36*(4), 37–66.

Forrester, J. W. (1961). *Industrial dynamics*. Portland, OR: Productivity Press.

French, M. L., & LaForge, R. L. (2006). Closed-loop supply chains in process industries: an empirical study of producer re-use issues. *Journal of Operations Management*, *24*(3), 271–286. doi:10.1016/j.jom.2004.07.012

Fuji Xerox. (2007). *Eco-efficiency and cleaner production case studies*. Retrieved October 9, 2007, from http://www.environment.gov.au/ settlements/ industry/corporate/ eecp/case-studies/ xerox-dfe.html

Fuji Xerox Australia Pty Limited. (2007). *Fuji Xerox innovation makes business and environmental sense* [Electronic Version]. Retrieved October 09, 2007, from http://www.fujixerox.com.au/ about/ media/ articles/546

Georgiadis, P., & Vlachos, D. (2004). The effect of environmental parameters on product recovery. *European Journal of Operational Research*, *157*(2), 449–464. doi:10.1016/S0377-2217(03)00203-0

Georgiadis, P., Vlachos, D., & Tagaras, G. (2006). The impact of product lifecycle on capacity planning of closed-loop supply chains with remanufacturing. *Production and Operations Management*, *15*(4), 514–527. doi:10.1111/j.1937-5956.2006.tb00160.x

Gray, C., & Charter, M. (2007). *Remanufacturing and product design: designing for the 7th generation*. Farnham, UK: The Centre for Sustainable Design, University College for the Creative Arts.

Guide, V. D. R. (2000). Production planning and control for remanufacturing: industry practice and research needs. *Journal of Operations Management*, *18*(4), 467–483. doi:10.1016/S0272-6963(00)00034-6

Guide, V. D. R., & Wassenhove, L. N. V. (2001). Managing product returns for remanufacturing. *Production and Operations Management*, *10*(2), 142–154. doi:10.1111/j.1937-5956.2001.tb00075.x

Gungor, A., & Grupta, S. M. (1999). Issues in environmentally conscious manufacturing and product recovery. *Computers & Industrial Engineering*, *36*(4), 811–853. doi:10.1016/S0360-8352(99)00167-9

Inderfurth, K. (2005). Impact of uncertainties on recovery behavior in a remanufacturing environment: a numerical analysis. *International Journal of Physical Distribution & Logistics Management*, *35*(5), 318–336. doi:10.1108/09600030510607328

Inderfurth, K., & van der Laan, E. (2001). Leadtime effects and policy improvement for stochastic inventory control with remanufacturing. *International Journal of Production Economics*, *71*(1-3), 381–390. doi:10.1016/S0925-5273(00)00135-3

Kiesmuller, G. P. (2003). A new approach for controlling a hybrid stochastic manufacturing/remanufacturing system with inventories and different leadtimes. *European Journal of Operational Research*, *147*(1), 62–71. doi:10.1016/S0377-2217(02)00351-X

Kiesmuller, G. P., & Minner, S. (2003). Simple expression for finding recovery system inventory control parameter values. *The Journal of the Operational Research Society*, *54*(1), 83–88. doi:10.1057/palgrave.jors.2601480

Kiesmuller, G. P., & van der Laan, E. A. (2001). An inventory model with dependent product demands and returns. *International Journal of Production Economics*, *72*(1), 73–87. doi:10.1016/S0925-5273(00)00080-3

Kirkwood, C. W. (1998). *System dynamics methods: a quick introduction.* Department of Supply Chain Management, Arizona State University.

Klausner, M., & Hendrickson, C. T. (2000). Reverse-logistics strategy for product take-back. *Interfaces*, *30*(3), 156–165. doi:10.1287/inte.30.3.156.11657

Kleber, R. (2006). *Dynamic inventory management in reverse logistics*. Heidelberg: Springer.

Koh, S.-G., Hwang, H., Sohn, K.-I., & Ko, C.-S. (2002). An optimal ordering and recovery policy for reusable items. *Computers & Industrial Engineering, 43*(1,2), 59–73.

KPMG. (2008). *MobileMuster: The official recycling program of the mobile phone industry 2007–08 Annual Report*. Sydney, NSW: Australian Mobile Telecommunications Association.

Kulwiec, R. (2006). Reverse logistics provides green benefits. *Target*, *22*(3), 11–20.

Mitra, S. (2007). Revenue management for remanufactured products. *Omega*, *35*(5), 553–562. doi:10.1016/j.omega.2005.10.003

Morecroft, J. (2007). *Strategic modelling and business dynamics: a feedback systems approach*. Chichester, UK: John Wiley & Sons.

ReturnBuy. (2000). *The new dynamics of returns: the profit, customer and business intelligence opportunities in returns*. Ashburn, Virginia: ReturnBuy.com.

Reverse Logistics Executive Council. (2007). *What is the size of reverse logistics in the US?* Retrieved 24 April, 2007, from http://www.rlec.org/ glossary.html

Rogers, D. S., & Tibben-Lembke, R. (2001). An examination of reverse logistics practices. *Journal of Business Logistics*, *22*(2), 129–148. doi:10.1002/j.2158-1592.2001.tb00007.x

Savaskan, R. C., Bhattacharya, S., & Wassenhove, L. N. V. (2004). Closed-loop supply chain models with product remanufacturing. *Management Science*, *50*(2), 239–252. doi:10.1287/mnsc.1030.0186

Sterman, J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. Burr Ridge, Illinois: McGraw-Hill.

Stock, J. R. (2001). The 7 deadly sins of reverse logistics. *Material Handling Management*, *56*, 5–11.

Tang, O., Grubbström, R. W., & Zanoni, S. (2007). Planned lead time determination in a make-to-order remanufacturing system. *International Journal of Production Economics*, *108*(1–2), 426–435. doi:10.1016/j.ijpe.2006.12.034

Teunter, R. H. (2001). Economic ordering quantities for recoverable item inventory systems. *Naval Research Logistics*, *48*(6), 484–495. doi:10.1002/nav.1030

Toktay, L. B., Wein, L. M., & Zenios, S. A. (2000). Inventory management of remanufacturable products. *Management Science*, *46*(11), 1412–1426. doi:10.1287/mnsc.46.11.1412.12082

van der Laan, E., Dekker, R., & Salomon, M. (1996). Product remanufacturing and disposal: a numerical comparison of alternative control strategies. *International Journal of Production Economics*, *45*(1–3), 489–498. doi:10.1016/0925-5273(95)00137-9

van der Laan, E., Dekker, R., Salomon, M., & Ridder, A. (1996). An (s, Q) inventory model with remanufacturing and disposal. *International Journal of Production Economics*, *46–47*, 339–350. doi:10.1016/0925-5273(95)00020-8

van der Laan, E., & Salomon, M. (1997). Production planning and inventory control with remanufacturing and disposal. *European Journal of Operational Research*, *102*(2), 264–278. doi:10.1016/S0377-2217(97)00108-2

van der Laan, E., Salomon, M., & Dekker, R. (1999). An investigation of lead-time effects in manufacturing/remanufacturing systems under simple PUSH and PULL control strategies. *European Journal of Operational Research*, *115*(1), 195–214. doi:10.1016/S0377-2217(98)00108-8

van der Laan, E., Salomon, M., Dekker, R., & Wassenhove, L. V. (1999). Inventory control in hybrid systems with remanufacturing. *Management Science*, *45*(5), 733–747. doi:10.1287/mnsc.45.5.733

van der Laan, E. A., & Teunter, R. H. (2006). Simple heuristics for push and pull remanufacturing policies. *European Journal of Operational Research*, *175*(2), 1084–1102. doi:10.1016/j.ejor.2005.06.030

Ventana Systems, Inc. (1999). *Vensim® PLE, Vensim® PLE Plus, Personal Learning Edition, User's Guide (Version 4)*. Harvard, USA: Ventana Systems, Inc.

Vlachos, D., Georgiadis, P., & Iakovou, E. (2007). A system dynamics model for dynamic capacity planning of remanufacturing in closed-loop supply chains. *Computers & Operations Research*, *34*(2), 367–394. doi:10.1016/j.cor.2005.03.005

Wells, P., & Seitz, M. (2005). Business models and closed-loop supply chains: a typology. *Supply Chain Management*, *10*(3–4), 249–251.

KEY TERMS AND DEFINITIONS

Closed Loop Supply Chain: Integration between a forward and reverse supply chain in reverse logistics processes.

Customer Relationship Management: Company strategies/policies for managing relationships with customers and organizing business processes.

Remanufacturing: A sustainable method applied particularly in closed loop supply chains to recapture value from product returns.

Reverse Logistics: A sustainable process applied in returns management in order to reuse, remanufacture, and recycle or properly dispose of product returns.

Sustainable Production: A production process that creates goods and services and is aimed at minimizing the usage of energy and natural resources.

System Dynamics: A modeling and simulation approach aimed at understanding the dynamic behavior of complex systems in order to analyze and solve complex problems with a focus on policy analysis and design.

Systems Modeling: Representation and construction of business systems through the use of models which correspond as much as possible to a meaningful concept based on real-world situations.

This work was previously published in Advanced Analytics for Green and Sustainable Economic Development: Supply Chain Models and Financial Technologies, edited by Zongwei Luo, pp. 157-186, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 20
A Production Planning Optimization Model for Maximizing Battery Manufacturing Profitability

Hesham K. Alfares
King Fahd University of Petroleum & Minerals, Saudi Arabia

ABSTRACT

This paper presents an integer programming (IP) model for production planning, which is used to maximize the profitability of battery manufacturing in a mid-size company. Battery production is a complicated multi-stage process. The formation stage, during which the batteries are filled with acid and charged with electricity, is considered to be the bottleneck of this process. The IP model maximizes the total profit of batteries produced in the formation stage, subject to limited manufacturing resources as well as time limitations and demand restrictions. The IP model is able to accommodate a large variety of battery models and sizes, and also different charging circuit capacities and speeds. The model is formulated and optimally solved using Microsoft Excel Solver. Compared to the current manual production planning approach, the optimum IP-generated production plans lead to an average increase of 12% in daily profits.

INTRODUCTION

Battery production requires a special manufacturing process that involves several distinctive steps such as chemical curing, acid filling, and electrical charging. Due to the unique characteristics of the battery manufacturing process, battery production planning has its own distinguishing features. In this paper, an integer programming model is presented for maximizing the profitability of battery production at the formation stage of manufacturing for the Middle East Battery Company (MEBCO) in Saudi Arabia. The model is formulated and applied to real production data, showing significant

DOI: 10.4018/978-1-4666-1945-6.ch020

advantages over the current manual production planning approach.

The Middle East Battery Company (MEBCO) is a joint venture between a group of Saudi industrialists and Johnson Control International (JCI). MEBCO produces AC Delco, JCI and Toyota branded maintenance-free batteries. As an equity partner in MEBCO, JCI has management responsibility for the plant operation. MEBCO plant, which is located in Dammam, Saudi Arabia, started commercial production in 1998. The production has been steadily growing and it has exceeded 2.5 million batteries since 2006. The plant produces batteries for all vehicle types, as well as marine, commercial and industrial applications. MEBCO's primary market is the Middle East, but it also has growing markets in Asia, Europe, and North America.

MEBCO's battery manufacturing plant is divided into 4 main areas; X-Met, Green Group, Formation, and Final Assembly. The Formation area's main function is the electrical charging of batteries for periods that range from 14.5 hours to 24 hours, depending on the type of battery. Because of the limited time per day, and also because of space limitations that restrict the number of charging circuits, the Formation area is considered to be the bottleneck of the battery production capacity.

According to the theory of constraints, the throughput rate of any multi-component system is controlled by the slowest component (called the constraint or the bottleneck). In order to maximize the productivity (profitability) of the battery production line, the company needs to optimize the production planning output in the formation stage. As shown in the following section, a variety of optimization and heuristic techniques have been used to solve battery production planning problems. In this paper, an integer programming (IP) model is used to maximize the profitability of the formation area. This IP model has been practically applied to optimally schedule the daily production of 74 battery models. The model considers different battery sizes, charging times, and daily demands, as well as limitations on times, capacities, and numbers of different types of charging circuits. Compared to manually generated production plans, the IP-optimized production plans increase daily profits by 12% on average.

The remaining sections of this paper are arranged as follows. First, previous literature on battery production planning is reviewed. Next, the manufacturing process, especially in the formation step, is described in more detail. Subsequently, the integer programming model for battery production planning is presented, and then the model's performance is compared to the current manual approach. Finally, results are analyzed, and conclusions are drawn.

LITERATURE REVIEW

Previous approaches to battery production planning include simulation, network flow models, EOQ-type techniques, mathematical programming, and various heuristics. Turnquist (1991) described a software system for battery production planning developed by Cornell University for the Battery Strategic Business Unit of Delco Remy. The PC-based system includes three main components: (1) a forecasting module, (2) a multi-plant product allocations module, and (3) a scheduling module. The production scheduling module employs a network flow model to dynamically balance inventory and overtime costs given limited capacities and fluctuating demands for each plant.

Yenradeea (1994) combines simulation, the optimized production technology (OPT), and simple scheduling rules to schedule a four-stage battery production line. The OPT production plans successfully minimize inventory while maximizing the throughput rate. Using simulation experiments, these plans are shown to outperform both the push and the pull policies. Khadem and Ali (2008) develop a simulation model to optimize the cost effectiveness of a car battery manufacturer.

The model is used to represent and analyze the dynamics of the battery assembly line, and also to make several recommendations for improving the line's cycle time, productivity, and quality.

Sharma (2003) integrated the principles of lean production with the practices of Six Sigma to improve daily work life in a battery company on a continuing basis. The advantages realized by the company included significant cost reduction, more efficient manufacturing process, and higher customer satisfaction. Al-Turki (2000) presented a heuristic multi-product batch scheduling algorithm to schedule battery batch production with the minimum set-up time. The effects of set-up time/frequency and batch size on the production performance were investigated in a real-life battery manufacturing company. César et al. (2007) described a case study implementation of intermediate stock management in the production process for automotive batteries, highlighting the main positive aspects and the deficiencies.

Huston (1996) used a three-stage process for production planning at the International Fuel Cell Corporation (IFC) in the US. First, a spread-sheet model determines daily production and workforce size. Next, an EOQ model determines optimal lot sizes. Finally, a process improvement model increases bottleneck production by reducing setup time and/or cycle time. Elimam and Udayabhanu (1999) presented two procedures for planning the production of lead-acid batteries: a simple MRP-based heuristic and a mixed integer programming optimization model. In a later paper, Elimam and Udayabhanu (2001) modeled battery production as a multi-product, multi-period, and multi-component supply chain process. The trade-offs in this process – including the costs of holding, ordering, and defects – were tested in a small battery manufacturing plant.

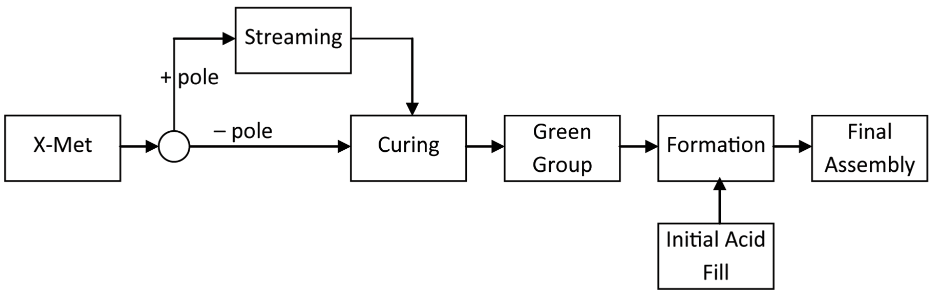
Dunstall and Mills (2002) presented a mixed-integer programming (MIP) model to optimally schedule cyclic production of batteries for Exide Technologies in Australia. A production schedule is developed for 42 periods in a cyclic workweek. The objective is to maximize throughout of a multi-product battery charging and finishing facility while maintaining a given product mix. Dunstall and Mills (2007) slightly modified their earlier MIP model and then used simulation experiments to evaluate its performance compared to two simple scheduling rules: shortest processing time (SPT), and random sequencing (RND). The simulation results confirmed that the optimal MIP schedules consistently yield higher throughputs than simple on-line dispatching rules.

The formation-stage production planning problem at MEBCO, which is described in detail later, does not involve any stochastic or nonlinear aspects. Although the problem has integer variables, it is fairly small and thus easy to solve. Therefore, optimization by integer linear programming is the most suitable approach for this problem. Compared to Dunstall and Mills (2002) MIP model, this paper presents a pure-integer programming model with substantially different decision variables, objective function, and constraints. Before introducing the model, the battery manufacturing process is described in the following section.

BATTERY MANUFACTURING PROCESS

The manufacturing space in the plant is divided into 4 main areas; X-Met, Green Group, Formation, and Final Assembly. The main processes in the X-Met area are lead strip manufacturing, oxide manufacturing, paste mixing, X-Met line, steaming, and curing. The Green Group main sections are encapsulation/collation, case preparation, battery assembly, cast-on-strap, side terminal weld, extrusion fusion weld, heat seal, and top terminal weld. The main processes in the Formation stage, which is described in more detail later in this section, are initial acid fill and formation. The Final Assembly's main processes are formation acid dump, final acid fill, final heat seal, labeling,

Figure 1. Flow chart of the manufacturing process



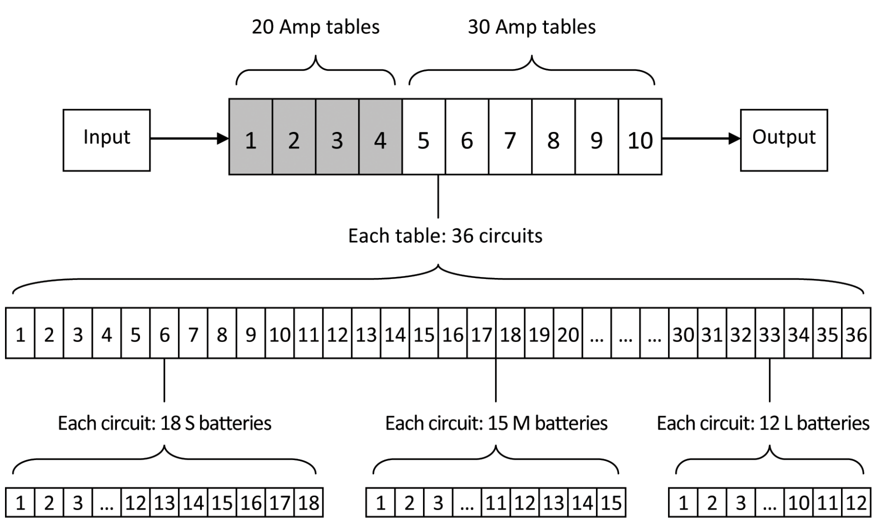
and packing. Figure 1 illustrates the flow of the manufacturing process.

The first operation in the formation area is filling the batteries with a lower-specific-gravity acid called the forming acid. This weak acid contains more water than acid to facilitate and promote the chemical conversion of the plates. This forming acid also contains a specific ingredient called formax. Controlling the amount of formax, acid gravity, and acid volume in the acid mixing process is very important in attaining proper battery formation and performance.

Once the Green Group batteries have been filled with forming acid, they are transferred by conveyors to the charging area, whose layout is shown in Figure 2. The charging area has 10 tables, 4 of which have 20-Ampere chargers while the remaining 6 have 30-Ampere chargers. On the charging tables, batteries are arranged in groups (known as circuits) to match the circuit output points of the chargers. Each charger has 36 output points (circuits), with up to 18 small-size batteries per circuit. However, due to space limitations, only 15 medium-size or 12 large-size batteries can be connected to each circuit. Within each circuit, the individual batteries are hooked together (positive of one battery attached to negative of next battery) by the use of charging straps.

Once the batteries are loaded on to tables in different circuits, they are subjected to a one-hour rest or pickling time. During this one-hour time frame, strap and circuit connections are com-

Figure 2. Layout of the charging tables in the formation area



pleted and the batteries are spaced apart to attain proper air flow cooling. Subsequently, they are given two steps of electrical charges for a specific duration and ampere rates. Between two steps, a rest time of 3 hours is provided. The electrical charges are applied in order to turn the green sulfated plate material into formed plates capable of storing and delivering electrical charge.

The electrical charging cycle of the batteries varies with the size of the battery. Smaller automotive batteries require less Ampere-hour (Amps-charger output multiplied by time on charge) charging than larger automotive batteries. Extra-large heavy duty batteries tend to have the longest charging cycles. Once the formation is completed, the batteries are unstrapped from each other and from the charging circuits. Finally, the batteries are off loaded from the charging rows onto conveyors that transfer the units to the finishing line. The steps of the formation process are depicted in Figure 3.

Sometimes during the peak season, the formation stage becomes a bottleneck that limits the battery production capacity because it involves many constraints and obstacles. These include long formation times, capacity constraints, lack of scheduling system, and inexact demand forecasting. In the following section, an integer programming optimization model for production planning is formulated for maximizing the productivity (profitability) of the formation stage.

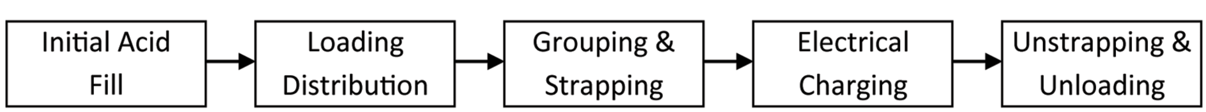
FORMATION CAPACITY OPTIMIZATION MODEL

In this section, we present the assumptions, decision variables, objective function, and constraints of an integer programming production planning model that is formulated to maximize the profitability of batteries produced from the formation area.

Assumptions of the Model

1. There are 10 tables for charging batteries. The number of tables is currently limited due to scarce space availability in the formation area.
2. There are six 30-Amps and four 20-Amps charging tables. This is probably a sub-optimum allocation of the 10 tables among the two types, because more battery models can be charged on 20-Amps circuits than on 30-Amps circuits.
3. Each table has 36 charging circuits. This number is fixed by the table size and the space requirements of the different battery models and charging circuits.
4. Each circuit consists of 18, 15 or 12 batteries, depending on battery size.
5. There are three battery sizes:
   - Size 1 (small): 18 batteries per circuit.
   - Size 2 (medium): 15 batteries per circuit.
   - Size 3 (large): 12 batteries per circuit.
6. In any charging circuit, we cannot have any mix among the three battery sizes.
7. Some batteries can be charged only on 20-Amps circuits, some others can be charged on 30-Amps, while the rest can be charged on both.
8. If a battery can be charged on either 20-Amps or 30-Amps circuits, then the charging time is shorter on the 30-Amps.

Figure 3. Detailed flow chart of the formation stage



9. The larger the battery size, the longer the charging time required.
10. The total number of battery models produced by the company is 74. Out of these, 35 models can be charged on 20-Amps circuits, 31 models can be charged on 30-Amps circuits, while 8 models can be charged on both types of circuits.
11. The maximum demand for each battery model is given per day, based on the overall weekly demand.

In reality, there are several other considerations that are not explicitly included in the model because they do not affect the model's output. For example, storage time constraints require batteries are not kept for more than 96 hours after formation. In addition, batteries taking longer times are not loaded when there is an urgent need for finishing certain quantities of other models. Finally, where demand is normal, batteries with longer charging duration will be put on charge first.

Given Parameters

d_i = demand of model i battery, i = 1,..., 74
p_i = profit per unit of model i battery, i = 1,..., 74
t_{ia} = formation time of model i battery if charged on a 20-Amps table
t_{ib} = formation time of model i battery if charged on a 30-Amps table
S_k = set of indices (model numbers) of batteries of size k, k = 1,..., 3

Decision Variables

X_i = number of model i batteries charged on 20-Amps tables per day, i = 1,..., 74
Y_i = number of model i batteries charged on 30-Amps tables per day, i = 1,..., 74
A_k = number of 20-Amps circuits assigned to batteries of size k, k = 1,..., 3
B_k = number of 30-Amps circuits assigned to batteries of size k, k = 1,..., 3

Objective Function

The objective of the integer programming model is to maximize the total daily profit of batteries processed in the formation stage of the battery manufacturing process:

$$\text{Maximize } Z = \sum_{i=1}^{74} p_i (X_i + Y_i) \quad (1)$$

Constraints

The maximum daily demand for each battery model cannot be exceeded:

$$X_i + Y_i \le d_i,\ i = 1,..., 74 \quad (2)$$

The number of batteries charged in the 20-Amps (30-Amps) tables is limited by the available respective circuit capacity:

$$A_1 + A_2 + A_3 \le 4(36) \quad (3)$$

$$B_1 + B_2 + B_3 \le 6(36) \quad (4)$$

For each of the three sizes, the total number of hours used for charging 20-Amps batteries cannot exceed the total available circuit-hours per day:

$$\sum_{i \in S_1} t_{ia} X_i \le 18(24) A_1 \quad (5)$$

$$\sum_{i \in S_2} t_{ia} X_i \le 15(24) A_2 \quad (6)$$

$$\sum_{i \in S_3} t_{ia} X_i \le 12(24) A_3 \quad (7)$$

For each of the three sizes, the total number of hours used for charging 30-Amps batteries cannot exceed the total available circuit-hours per day:

$$\sum_{i \in S_1} t_{ib} Y_i \le 18(24) B_1 \quad (8)$$

$$\sum_{i \in S_2} t_{ib} Y_i \le 15(24) B_2 \quad (9)$$

$$\sum_{i \in S_3} t_{ib} Y_i \le 12(24) B_3 \quad (10)$$

X_i, Y_i, A_k, $B_k \ge 0$ and integer, $i = 1,..., 74$, $k = 1,..., 3$ (11)

MODEL APPLICATION AT MEBCO

All of the data required for the model were collected from the MEBCO Product Standards Manual for all 74 battery models produced by MEBCO. Some of these models can be charged either on 20-Amps or on 30-Amps charging circuits. Table 1 shows standard data (formation time, charge table type, and battery size) for a sample of 15 battery models. In addition, current unit profits and typical daily demands are shown for these 15 battery models.

For each work day, the above integer programming model was solved for different daily demands d_i, $i = 1,..., 74$, using the Solver optimization tool in Microsoft Excel. Compared to the current manual schedules, the optimum solutions found by the model produced significantly higher daily profits. The increase in daily profit ranged from \$3,500 to \$6,200, and averaged around \$4,600. In relative terms, the increase in daily profit ranged from 8% to 16%, and averaged around 12%. Table 2 compares the daily profits of the model's optimum solution and the company's manual solutions for 5 typical work days.

Table 1. Typical data for a sample of 15 battery models

Model	Formation time (Hr)	Table type	Battery size	Profit ($)	Daily demand
1	17.5	20 Amps	1	3.52	748
4	21	20 Amps	1	4.59	218
	17	30 Amps			
5	24	30 Amps	2	6.72	4
20	22.5	30 Amps	3	6.72	180
21	23.5	20 Amps	2	5.28	75
22	23.5	20 Amps	2	5.28	225
28	24	20 Amps	1	5.23	0
29	24	30 Amps	2	11.12	0
30	21	30 Amps	3	7.15	72
31	16	20 Amps	1	3.52	192
	15.5	30 Amps			
32	23	30 Amps	3	7.87	0
33	15.5	20 Amps	1	3.52	256
	16	30 Amps			
34	24	30 Amps	3	10.40	9
35	18	20 Amps	1	3.04	0
36	16.5	20 Amps	1	3.04	0
37	20.5	30 Amps	2	9.20	144

Table 2. Optimal solution profit vs. company schedules' profits for five work days

Day	Company profit ($)	Optimum Profit ($)	% Profit Increase
1	37,387	43,200	**15.55**
2	41,605	46,160	**10.95**
3	32,803	37,099	**13.10**
4	41,369	45,077	**8.97**
5	35,973	39,767	**10.54**

CONCLUSION

An optimum production planning model has been formulated and used to maximize the profitability of battery production for a real-life battery manufacturer. The integer programming model is specifically designed to determine the production plan at the formation stage (i.e., quantities of each battery model allocated to each type of charging table). The constraints of the model include demand for each battery model, availability of charging tables, and charging time limitations.

The IP model has several advantages. First, the model produced significant improvements (averaging 12%) in daily profit. Second, the model is flexible to changes in input parameters such as daily demands, or to changes in decision variables (e.g., new battery models). Another advantage is the use of standard Microsoft Excel Solver, which makes the model a very convenient low-cost tool to find the optimal battery production plan under a variety of scenarios. Limitations of the model are imposed by the given assumptions and constraints, reflecting the company's current practice. These include space and time constraints and the number of charging tables of each type.

Several opportunities for future research could be pursued in order to improve performance further. First, daily demands for different battery types could be estimated more carefully in order to take full advantage of the model. Second, the optimum number of 20-Amps and 30-Amps tables could be determined under space and budget constraints, subject to varying daily demands for individual battery types. Third, mixing different battery sizes on the same charging circuit could be allowed to increase the flexibility to meet demands. Finally, the IP model could be combined with facility layout techniques in order to maximize the space utilization and increase the number of charging tables.

REFERENCES

Al-Turki, Y. A. Y. (2000, April 1-4). Multi-product batch scheduling for lead-acid battery assembly line. In *Proceedings of the 11th Annual Production and Operations Management Society Conference*, San Antonio, TX.

César, P., Rodrigues, C., & Oliveira, O. J. (2007, May 4-7). Intermediate stock management in the production process for automotive batteries: a case study. In *Proceedings of the 18th Annual Production and Operations Management Society Conference*, Dallas, TX.

Dunstall, S., & Mills, G. (2002, February 11-15). Scheduling the charging of batteries. In *Proceedings of the Mathematics in Industry Study Group*, Adelaide, Australia (pp. 87-105).

Dunstall, S., & Mills, G. (2007). Robustness of cyclic schedules for the charging of batteries. *ANZIAM Journal*, *48*(4), 475–492. doi:10.1017/S1446181100003163

Elimam, A. A., & Udayabhanu, V. (1999, May 23-24). Production planning for lead-acid batteries. In *Proceedings of the IIE Industrial Engineering Research Conference*, Phoenix, AZ.

Elimam, A. A., & Udayabhanu, V. (2001, August 6-8). Multi-level supply chain in lead-acid battery manufacturing. In *Proceedings of the First International Conference on Logistics and Supply Chain Management*, Coimbatore, India (pp. 99-103).

Huston, T. W. (1996). *Production system design and cycle time reduction in a fuel cell manufacturing operation* (Unpublished master's thesis). MIT, Cambridge, MA.

Khadem, M., & Ali, A. (2008, November 16-18). Modeling and simulation for car battery manufacturing for cost effectiveness. In *Proceeding of the Intelligent Systems and Control Symposia: Computational Biology and Bioinformatics*, Orlando, FL.

Middle East Battery Company. (2006). *MEBCO product standards manual*. Dammam, Saudi Arabia: Middle East Battery Company.

Sharma, U. (2003). Implementing lean principles with the Six Sigma advantage: How a battery company realized significant improvement. *Journal of Organizational Excellence*, *22*(3), 43–52. doi:10.1002/npr.10078

Turnquist, M. A. (1991, October 27-31). University/industry cooperation on an integrated production planning software system. In *Proceedings of the Conference on Technology Management: The New International Language*, Portland, OR (pp. 456-459).

Yenradeea, P. (1994). Application of optimized production technology in a capacity constrained flow shop: a case study in a battery factory. *Computers & Industrial Engineering*, *27*(1-4), 217–220. doi:10.1016/0360-8352(94)90274-7

This work was previously published in International Journal of Applied Industrial Engineering (IJAIE), Volume 1, Issue 2, edited by Rajbir Singh Bhatti, pp. 55-63, copyright 2012 by IGI Publishing (an imprint of IGI Global).

Chapter 21
Multi-Objective Optimization of Manufacturing Processes Using Evolutionary Algorithms

M. Kanthababu
Anna University, India

ABSTRACT

Recently evolutionary algorithms have created more interest among researchers and manufacturing engineers for solving multiple-objective problems. The objective of this chapter is to give readers a comprehensive understanding and also to give a better insight into the applications of solving multi-objective problems using evolutionary algorithms for manufacturing processes. The most important feature of evolutionary algorithms is that it can successfully find globally optimal solutions without getting restricted to local optima. This chapter introduces the reader with the basic concepts of single-objective optimization, multi-objective optimization, as well as evolutionary algorithms, and also gives an overview of its salient features. Some of the evolutionary algorithms widely used by researchers for solving multiple objectives have been presented and compared. Among the evolutionary algorithms, the Non-dominated Sorting Genetic Algorithm (NSGA) and Non-dominated Sorting Genetic Algorithm-II (NSGA-II) have emerged as most efficient algorithms for solving multi-objective problems in manufacturing processes. The NSGA method applied to a complex manufacturing process, namely plateau honing process, considering multiple objectives, has been detailed with a case study. The chapter concludes by suggesting implementation of evolutionary algorithms in different research areas which hold promise for future applications.

1. INTRODUCTION

Optimization refers to determining one or more feasible solutions from a set of available alternatives which may correspond to the maximum or minimum value of one or more objective functions. It implies choosing values for a variable from a set of real numbers such that the desired function takes one of the values. Finding an optimal solution for a given problem, which involves only one objective function is called single-objective optimization. Single-objective optimization generally attempts to find one good solution that is better than any

DOI: 10.4018/978-1-4666-1945-6.ch021

other feasible solutions in a set or search space. The input process parameters corresponding to such solution are called optimal parameters. For a single-objective problem, varieties of optimization methods ranging from classical to search methods have been used. Researchers have used the single-objective optimization techniques in several fields such as routing, scheduling, salesman, image processing, engineering design, parameter fitting, computer game playing, transportation, etc. (Goldberg, 1989; Srinivas and Deb, 1995).

A single-objective optimization involves maximizing or minimizing a chosen objective function without considering its effects on any other objective functions. Consider a hypothetical case, where the customer wants a product with the best possible surface finish irrespective of the cost/material removal rate. In such a scenario, the manufacturing engineer can choose the input process parameters that can yield the fine surface finish without considering the productivity or costs. On the other hand, if the desired outputs are more than one, for example when a customer demands for fine surface finish at lower cost, then it is considered as a multi-objective optimization problem. Therefore, the manufacturing engineer has to optimize two conflicting objectives i.e., fine surface finish and high material removal rate. If the manufacturing engineer chooses the first option based on the single-objective optimization considering only fine surface finish, then the productivity will suffer. Since, the fine surface finish results from lower material removal rate. If the manufacturing engineer chooses the input process parameters which provides only higher material removal rate, then it is very difficult to achieve the fine surface finish. Now the question arises, which is the best possible solution/solutions to achieve both fine surface finish and higher material removal rate together. In such a scenario, there is no single optimum solution or they cannot be solved through single-objective optimization techniques. Therefore, such problem is considered under the preview of multi-objective optimization.

Researchers made several attempts to solve multi-objective problems through different methods such as classical optimization, weighted sum, goal programming, min-max, etc. But these methods are not able to find multiple solutions in a single run. These methods change the multi-objective problem into a single-objective problem with the corresponding weights based on their relative importance and also suffer from a drawback that the decision maker must have a thorough knowledge of ranking of objective functions. These methods also fail when the objective functions become discontinuous. In contrast, evolutionary algorithms have been proved successfully for solving multi-objective problems because they operate with a population of individuals and also well suited to search for multiple solutions simultaneously. It is well proven that evolutionary algorithms can solve several conflicting objectives and also able to approximate the optimal solutions in a single run. Therefore, evolutionary algorithms have attracted a lot of research during the last two decades and it is still one of the hottest research areas in the field of evolutionary computation (Fonseca and Fleming, 1995; Sbalzariniy et al., 2000; Deb, 2001; Deb et al., 2002; Zitzler et al., 2004; Konak et al., 2006; Zhou et al., 2011).

In the view of the above, the following sections present the basic principles of multi-objective optimization as well as evolutionary algorithms and the working mechanisms of one of the efficient evolutionary algorithm has been detailed with a case study for ease in understanding. Finally the chapter is concluded with some potential directions for future research.

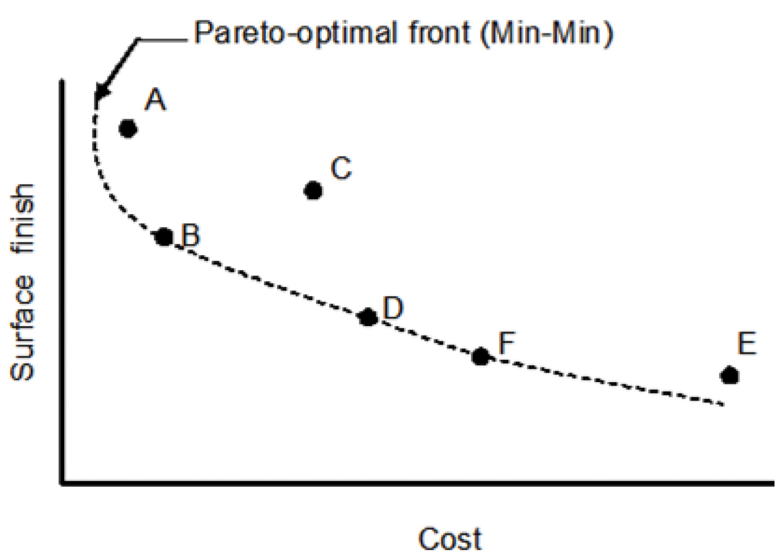
2. MULTI-OBJECTIVE OPTIMIZATION

In a multi-objective optimization problem there are more than one objective functions. The objective functions are often conflicting to each other and there is no single combination of optimal parameters, which can provide the best performance. Therefore, multi-objective optimization results in set of solutions in the search space. Among the set of solutions, no solution can be said to be inferior or superior over the other solutions in the set. Such solutions are known as Pareto-optimal solutions or Pareto-set or non-dominated solutions. The rest are called dominated solutions. The dominated solutions are not important for us. Therefore, in the multi-objective optimization, Pareto-optimal solutions are obtained by capturing number of feasible solutions simultaneously. Pareto-optimal solutions are often preferred than single solution because they are found to be practical while considering real life problems, since the final solution of the decision maker is always a trade-off. Almost, all real world problems can be accurately modelled by considering it as a multi-objective problem. Let us see Pareto-optimal solutions with a typical example as shown in Figure 1.

The example shown in Figure 1 has two objectives namely cost and surface finish, both of which are to be minimized (Min-Min) for productivity. In Figure 1, the point A represents a solution which incurs minimal cost but results in higher surface finish. On the other hand, the point E represents a solution which is costly, but it result in lower surface finish comparatively to point A. If both objectives i.e cost and surface finish are important goals and have to be considered simultaneously by the manufacturing engineer, one cannot really say whether solution A is better than solution E, or vice versa. i.e. Solution A is better than other in one objective (cost), but it is worse in the other objective (surface finish). But, there exists many solutions like B, D and F, which belong to the Pareto-optimal solutions. The Pareto-optimal solutions are shown as Pareto-optimal front marked as dashed line in Figure 1. From Figure 1, we can also observe that there exist non-Pareto-optimal solutions or dominated solution like point A, C and E. If we compare solution A with solution C, it is difficult to conclude which one is better than the other considering both the objectives. This is because there is another solution namely B in the search space, which is better than the solutions A and C.

The location of the Pareto-optimal front of the Pareto-optimal solutions in the search space

Figure 1. Pareto-optimal solutions

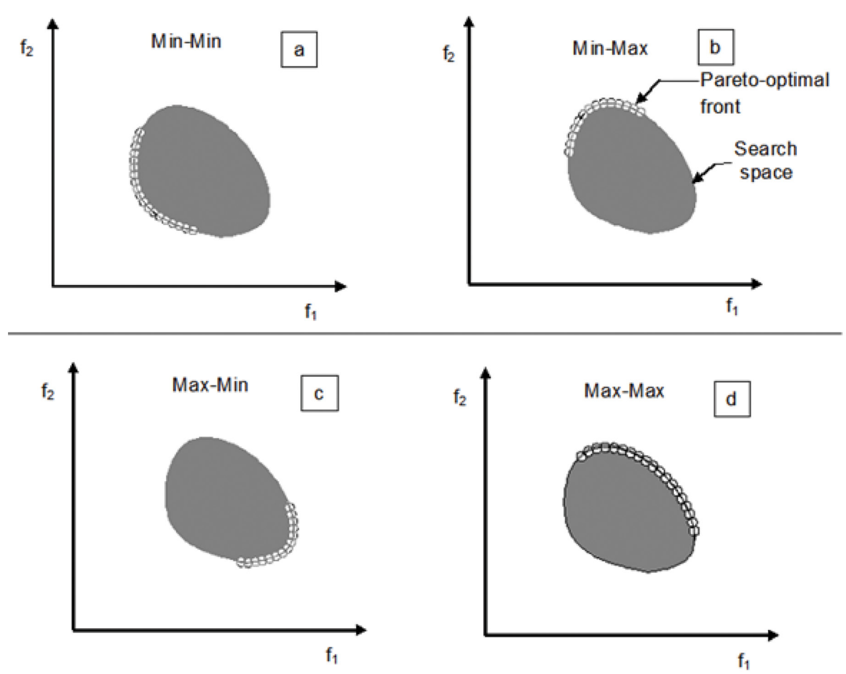


is based on the constraints of the objective functions, whether the objective functions have to be minimized or vice versa or opposite in nature (Figure 2). Consider the problem in Figure 2a, in which both objective functions f_1 and f_2 have to be minimized (Min-Min). The Pareto-optimal front (Pareto-optimal solutions) is shown with smaller circles in the search space. The search space is shown as shaded area in the Figure 2. The Pareto-optimal solutions in the front is a set of solution indicated that no one solution dominates the others and each solution is equally good for the given optimization problem and constraints. Figure 2b shows the Pareto-optimal front for a problem involving minimizing f_1 and maximizing f_2 (Min-Max) and Figure 2c shows the Pareto-optimal front corresponds to the problem of maximizing f_1 and minimizing f_2 (Max-Min). Figure 2d corresponds to the Pareto-optimal front for maximization of both objectives f_1 and f_2 (Max-Max).

The ultimate goal of the multi-objective algorithms for solving multi-objective problems is to identify solutions with in the Pareto-optimal solutions or Pareto-optimal front. However, it is practically difficult to identify the entire Pareto-optimal solutions due to its size or computational infeasibility. Therefore, a practical approach is to investigate a set of solutions (the best-known Pareto-set) that represent the Pareto-optimal set. Therefore, any multi-objective algorithms should achieve the following three important conflicting goals (Konak et al., 2006):

1. The best known Pareto-front should be as close as possible to the true Pareto-front. Ideally, the best known Pareto-set should be a subset of the Pareto-optimal set.
2. Solutions in the best known Pareto-set should be uniformly distributed and diverse over the Pareto-front in order to provide the decision maker a true picture of trade-offs.
3. The best known Pareto-front should capture the whole spectrum of the Pareto-front. This requires investigating solutions at the extreme ends of the objective function space.

Figure 2. Pareto-optimal front for different combinations of maximization and minimization problems



A typical multi-objective optimization problem consists of a number of objectives and is associated with a number of inequality and equality constraints. Mathematically, the multi-objective optimization problem can be expressed as follows:

Minimize / Maximize

f_i (x) i = 1, 2,....., N (1)

Subject to

$g_i(x) \leq 0$ j = 1, 2,....., J

$h_k(x)=0$ k=1, 2, ….., K

The parameter x is a p dimensional vector having p design or decision variables. Solutions to a multi-objective optimization problem are mathematically expressed in terms of Pareto-optimal solutions or non-dominated or superior points. In a minimization problem, a vector $x^{(1)}$ is less than another vector $x^{(2)}$, ($x^{(1)} < x^{(2)}$), when no value of $x^{(2)}$ is less than $x^{(1)}$ and at least one value of $x^{(2)}$ is strictly greater than $x^{(1)}$. If $x^{(1)}$ is partially less than $x^{(2)}$, we say that the solution $x^{(1)}$ dominates $x^{(2)}$ or the solution $x^{(2)}$ is inferior to the $x^{(1)}$. Any member of such vectors which is not dominated by any other member is said to be non-dominated or non-inferior. Similarly if the objective is to maximize a function, we define a dominated point if the corresponding component is greater than that of a non-dominated point (Rao, 1984).

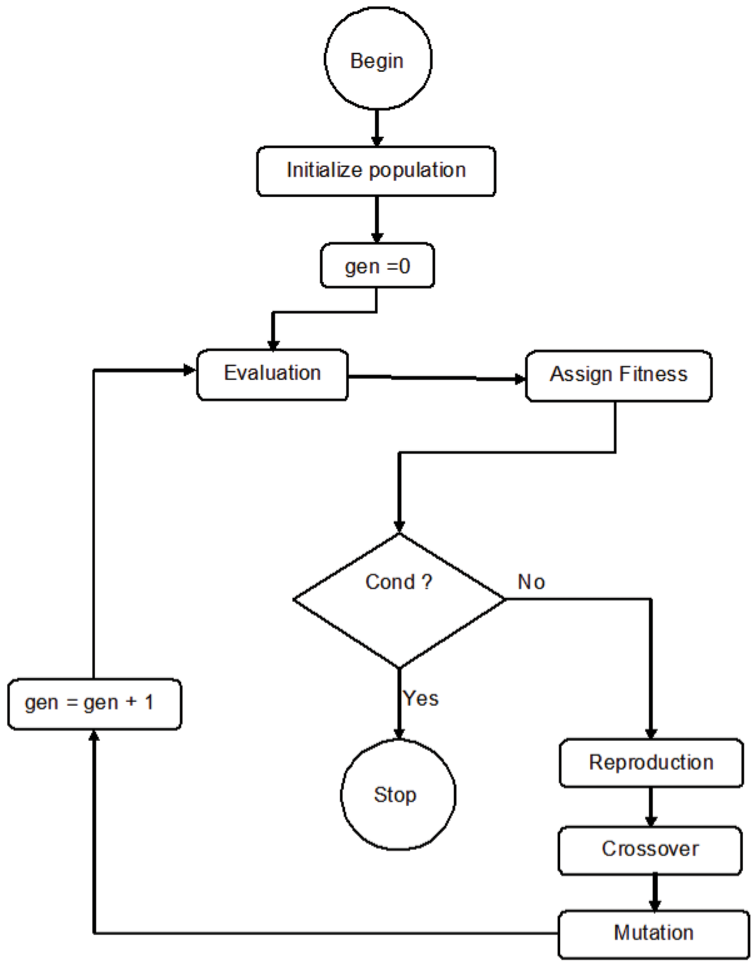
Mathematically, an optimization algorithm should terminate if any one of the Pareto-optimal solutions is obtained. But in practice, since there could be a number of Pareto-optimal solutions and the suitability of one solution depends on a number of factors. Hence, a multi-objective optimization algorithm must successfully find a globally optimal solution without getting restricted to local optima. Therefore, researchers are increasingly focussing on solving the multi-objective optimization problems using evolutionary algorithms due to its effectiveness and efficiency (Deb, 1995; Deb, 1999; Deb, 2001). Considering the importance of evolutionary algorithms for solving multiple objectives, the following are discussed in detail.

3. EVOLUTIONARY ALGORITHMS

The following sections deals with the different types of evolutionary algorithms starting with Genetic Algorithm (GA). GA is one of the widely used evolutionary algorithms. The procedure of a GA is given in Figure 3.

The concept of GA was developed by Holland and his colleagues in the 1960s (Goldberg, 1989). GA is a search and optimization tool developed by mimicking the evolutionary principles and chromosomal processing in natural genetics based on the concept of Darwinian theory of evolution. According to this theory, in nature, weak and unfit species within their environment are faced with extinction by natural selection. The strong ones have greater opportunity to pass their genes to future generations via. reproduction. In the long run, species carrying the correct combination in their genes become dominant in their population. Sometimes, during the slow process of evolution, random changes may occur in genes. Random changes provide additional advantages for survival (i.e) new species evolve from the old ones. Unsuccessful changes are eliminated by natural selection. The basics and working principles of GAs are detailed by Deb (1995) and Deb (2001). GA and its variations such as ant colony optimization, particle swarm optimization, simulated annealing, etc. are used by researchers for optimization. However, GA is mostly limited to single-objective optimisation. For addressing multi-objective optimization, several multi-objective evolutionary algorithms are formulated by different researchers to solve multi-objective problems (Fonseca and Fleming, 1995; Sbalzariniy et al., 2000; Deb 2001; Deb et al., 2002; Zitzler et al., 2004; Konak et al., 2006; Zhou et al., 2011).

Figure 3. Flowchart of GA (Deb, 2001)



3.1 Multi-Objective Optimization Algorithms

Multi-objective optimization algorithms differ from single objective optimization algorithms in terms of their fitness assignment procedure, elitism or diversification approaches and it can overcome the shortcomings encountered in single-objective optimization algorithms. The first multi-objective evolutionary algorithm called Vector Evaluated GA (or VEGA) was proposed by Schaffer (1985). Afterwards, several multi-objective evolutionary algorithms were developed including Multi-objective Genetic Algorithm (MOGA), Niched Pareto Genetic Algorithm (NPGA), Weight-based Genetic Algorithm (WBGA), Random Weighted Genetic Algorithm (RWGA), Non-dominated Sorting Genetic Algorithm (NSGA), Strength Pareto Evolutionary Algorithm (SPEA), improved SPEA (SPEA2), Pareto-Archived Evolution Strategy (PAES), Pareto Envelope-based Selection Algorithm (PESA), Region-based Selection in Evolutionary Multi-objective Optimization (PESA-II), Non-dominated Sorting Genetic Algorithm-II (NSGA-II), Multi-objective Evolutionary Algorithm (MEA), Micro-GA, Rank-Density Based Genetic Algorithm (RDGA), Dynamic Multi-objective Evolutionary Algorithm (DMOEA), etc (Fonseca and Fleming, 1993; Deb, 2001; Valenzuela, 2002; Konak et al., 2006). The summary of applications of multi-objective evolutionary algorithms for real world problems has been detailed by Zhoua et al. (2011). Although several algorithms exist, many researchers applied strategies from various multi-objective GAs and their performances were tested in several studies (Zitzler et al., 2000; Konak et al., 2006). The advantages and disadvantages of different evolutionary algorithms are briefly summarized and presented in Table 1.

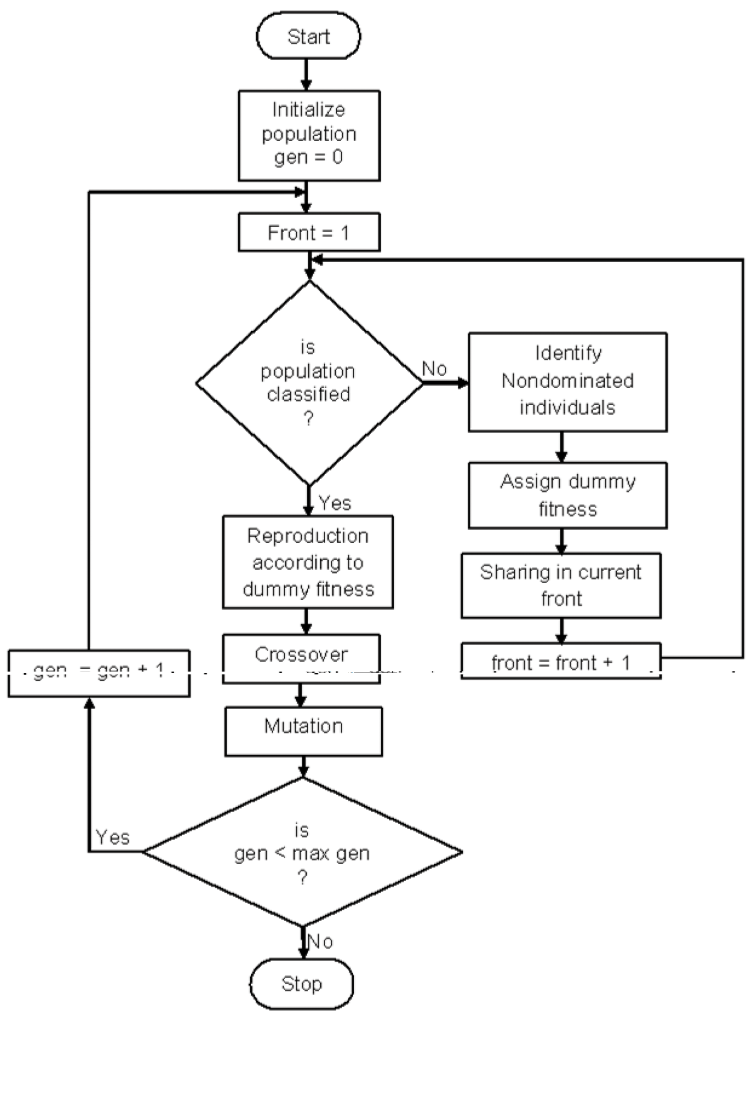
Though many evolutionary algorithms exist, only few researchers made an attempt to optimise multiple objectives using evolutionary algorithms especially in manufacturing process. Literature related to the application of evolutionary algorithms in manufacturing processes is briefly presented here. Kuriakose and Shunmugam (2005) made an attempt to optimise conflicting multiple objectives such as maximizing cutting velocity and minimising surface finish in wire electric discharge machining using NSGA. Sardiňas et al. (2006) made an attempt to optimise two mutually conflicting multiple objectives such as maximizing metal removal rate (to increase productivity) and minimising the delamination factor (improving surface quality) for drilling composite materials. Wang et al. (2006) made an attempt to optimise the multi-pass milling process considering multiple objectives such as minimising machining time as well as minimising production cost using parallel genetic simulated annealing, in which the fitness assignment is used based on

Table 1. Comparisons of evolutionary algorithms (Konak et al., 2006)

S. No	Multi-objective evolutionary algorithm	Advantages	Disadvantages
1	Vector Evaluated Genetic Algorithm (VEGA)	Straightforward for implementation	Tend to converge to the extreme for each objective
2	Multi objective Genetic Algorithm (MOGA)	Simple extension of single objective GA	Slow convergence
3	Weight Based Genetic Algorithm (WBGA)	Simple extension of single objective GA	Difficulties in non-convex objective function space
4	Niched Pareto Genetic Algorithm (NPGA)	Very simple selection process with tournament selection	Requires extra parameters for tournament selection
5	Random Weighted genetic Algorithm (RWGA)	Efficient and easy to implement	Difficulties in non-convex objective function space
6	Pareto Envelope based Selection Algorithm (PESA)	Easy to implement and computationally efficient	However, performance is depends on the cell sizes
7	Pareto Archived Selection Strategy (PAES)	Easy to implement and computationally efficient	Not a population based approach and the performance is depends on the cell sizes
8	Non-dominated Sorting Genetic Algorithm (NSGA)	Fast convergence	Problems related to niche size parameter
9	Non-dominated Sorting Genetic Algorithm-II (NSGA-II)	Single parameter, well tested and efficient	Crowding distance works in objective space only
10	Strength Pareto Evolutionary Algorithm (SPEA)	Well tested and no parameter for clustering	Complex clustering algorithm
11	SPEA II	Ensures extreme points are preserved	Computationally expensive fitness and density calculations
12	Rank Density based Genetic Algorithm (RDGA)	Dynamic cell update and robust with respect to number of objectives	Difficulty to implement
13	Dynamic Multi-objective Genetic Algorithm (DMOEA)	Efficient techniques to update cell densities and adaptive approaches to set GA parameters	Difficulty to implement

the concept of NSGA. Mandal et al. (2007) made an attempt to optimize multi-objectives such as maximizing metal removal rate and minimizing tool wear (to maintain accuracy) in an electrical discharge machining process using NSGA-II. Yang and Natarajan (2010) made an attempt to optimize conflicting multi-objectives (i.e) to minimise tool wear and maximize metal removal rate in turning process by using multi-objective differential evolution (MODE) algorithm and NSGA-II. From the literature, it is observed that among the evolutionary algorithms, NSGA and NSGA-II methods are found to be advantageous than other algorithms, especially for selecting optimal parameters for manufacturing processes. It is also observed that NSGA and NSGA-II methods resulted in wider solutions for the manufacturing engineers to select the appropriate input process parameters in various manufacturing processes. They are also found to be stable and have uniform reproductive potential across non-dominated individuals (Srinivas and Deb, 1995; Nandasana et al., 2003; Konak et al., 2006; Yang and Natarajan, 2010). Hence, the principles of NSGA and NSGA-II are briefly presented in the following sections.

Figure 4. Flowchart of NSGA (Srinivas and Deb, 1995)



3.2 Non-Dominated Sorting Genetic Algorithm (NSGA)

Figure 4 shows a flowchart of NSGA.

NSGA was developed by Srinivas and Deb (1995) based on the limitations of VEGA. It is similar to a simple GA except the classification of non-dominated fronts and the sharing operation. Thus the NSGA varies from a simple GA only in the way the selection operator is used. The crossover and mutation operators remain as usual. In NSGA, ranking selection method is used to emphasize current non-dominated points and a niching method is used to maintain diversity in the population. Before the ranking selection is performed, the population is ranked on the basis of an individual's non-dominated. The non-dominated individuals are assumed to constitute the first non-dominated front in the population and assigned a large dummy fitness value. The same fitness value is assigned to give an equal reproductive potential to all these non-dominated individuals. In order to maintain diversity in the population, these classified individuals are then shared with their dummy fitness values. Sharing is achieved by performing selection operation using degraded fitness values which are obtained by a quantity proportional to the number of individuals around it. This causes multiple optimal points to co-exist in the population. After sharing, these non-dominated individuals for the second non-dominated front are obtained. These new set of points are then assigned a new dummy fitness value which is kept smaller than the minimum shared dummy fitness of the previous front. This process is continued until the entire population is classified into several fronts. The population is then reproduced according to the dummy fitness values. Since individuals in the first front have the maximum fitness value, they always get more copies than the rest of population. This was intended to search for non-dominated regions or Pareto-optimal fronts. This results in quick convergence of the population towards non-dominated regions or Pareto-optimal fronts and sharing helps to distribute it over this region. By emphasizing non-dominated points, NSGA is actually processing the Pareto-optimal regions. The efficiency of NSGA lies in the way multiple objectives are reduced to a dummy fitness function using non-dominated sorting procedure. Another aspect of NSGA is that it can practically compute any number of objectives. Both minimization and maximization problems can also be handled by this algorithm. The only place of change required for above two cases is in the way the non-dominated points are identified. Further details on NSGA can be obtained from Srinivas and Deb (1995) and Deb (2001).

3.3 Honing Process

The following sections deals with working principles of honing process and its surface topography. The importance of bearing area parameters and its influences on the performance has been discussed. Typical application of one of the most efficient evolutionary algorithm namely NSGA to plateau honing process for solving multiple objectives has been described.

3.3.1 Basics of Honing Process

Honing is a type of super-finishing process generally applied to internal cylindrical surfaces, but sometimes on few external surfaces also as a finishing operation. Honing operation is carried out with bonded abrasives sticks/stones (known as honing tool) fitted on the honing shoe of a mandrel (honing head). The honing tool is subjected to simultaneous action of rotation, reciprocation (oscillation) and expansion in radial direction resulting in grooves with crosshatch lay surface texture (Davis, 1989).

In the case of manufacturing of cylinder liners, the cylinder liners are subjected to rough honing, finish honing and plateau honing operations. Rough honing is performed on the bored surface with coarse abrasives to control the dimension of the cylinder liner. Finish honing is performed with medium size abrasives to improve the finish. Finally plateau honing is performed with fine size abrasives to remove only the peaks on the finish honed surface (in the order of few microns) to achieve desired surface topography. The valleys of the profiles of the cylinder liner surface obtained after finish honing is remains and untouched. Hence, the surface generated in the final component has texture from two processes (finish and plateau honing). Plateau honing is carried out in order to generate surface that resemble run-in surfaces. Therefore, the final liner surface texture has proper plateau portion at the top to resist wear and grooves at desired honing angle to retain lubricant during engine operation. However, the use of the traditional surface parameters such as R_a, in the analysis of multi-process texture is found to be not adequate to characterize the surface texture of the cylinder liners. Therefore, the engine manufacturers and researchers have established alternative procedures to accept the cylinder liners based on the bearing area parameters in the final finished component (after plateau honing) (Pawlus, 1994; Pawlus, 1997; Feng et al., 2003; Kanthababu et al., 2009a; Kanthababu et al., 2009b; Kanthababu et al., 2010).

3.3.2 Bearing Area Parameters

The surface topography of honed cylinder liner is characterized by the bearing area parameters using the bearing area curve. The bearing area curve is obtained by double-step filtering in accordance with ISO 13565-1:1996 from the surface roughness measuring instrument. The procedure is briefly presented here. Initially, a mean line is obtained by filtering the primary profile with phase-corrected Gaussian filter (ISO 11562:1996) with a selected cut-off (say 0.8 mm) (Figure 5a). During the double-step filtering, the entire valley portion which lies below the mean line obtained from the phase-corrected Gaussian filter are suppressed and the same filter is used once again on this suppressed profile (Figure 5b (i, ii and iii)). The mean line obtained on the suppressed profile is the final mean line, which is taken as final reference line. This reference line is transferred to the original primary profile and subtracted to get the roughness profile (Figure 5b (iv)). Then the bearing area curve is obtained from this roughness profile (Figure 6).

Figure 6 shows a typical roughness profile and its bearing area curve along with methods for their evaluation. Though definition of R_a is clear from Figure 6, a brief description of the bearing area parameters is given here to understand their nature. The bearing area parameters consist of reduced peak height (R_{pk}), depth of the core roughness (R_k), reduced valley height (R_{vk}), material portion corresponding to the upper limit position of the roughness core (M_{r1}) and material portion corresponding to the lower limit position of the roughness core (M_{r2}). A line segment representing 40% material ratio is moved over the curve starting from extreme left point. The slope of the line, when it matches first with the curve, is taken as slope of an equivalent straight line. This equivalent straight line is extended to meet 0% and 100% lines and difference in height denotes R_k, namely core roughness. Horizontal lines drawn through these intersection points at 0% and 100% lines meet the curve at M_{r1} and M_{r2} respectively. Points corresponding to R_{pk} and R_{vk} are determined such that shaded triangular areas represent area of the curve above and below the core roughness respectively. Further details can be obtained from ISO 13565-2:1996. It may be noted that the equivalent straight line influences the bearing area parameters (Pawlus, 1998).

Among the bearing area parameters, M_{r2} and $R_{k,}$ have been assigned with higher weightage of 50% and 30% respectively based on their impact of engine performance, while $R_{vk,}$ $R_{pk,}$ and M_{r1} are assigned with 15%, 4%, and 1% weightage respectively (Feng et al., 2003). A plateau honed surface with M_{r2} and R_k values within the specified range

Figure 5. Schematic of filtering process using phase-corrected filter

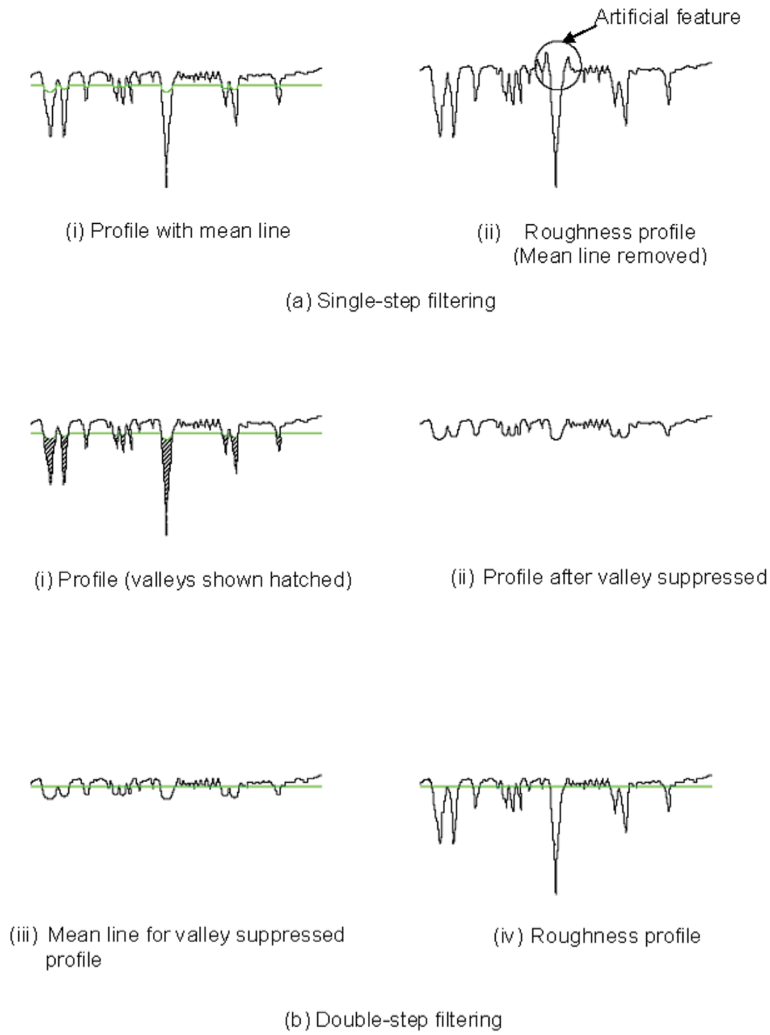
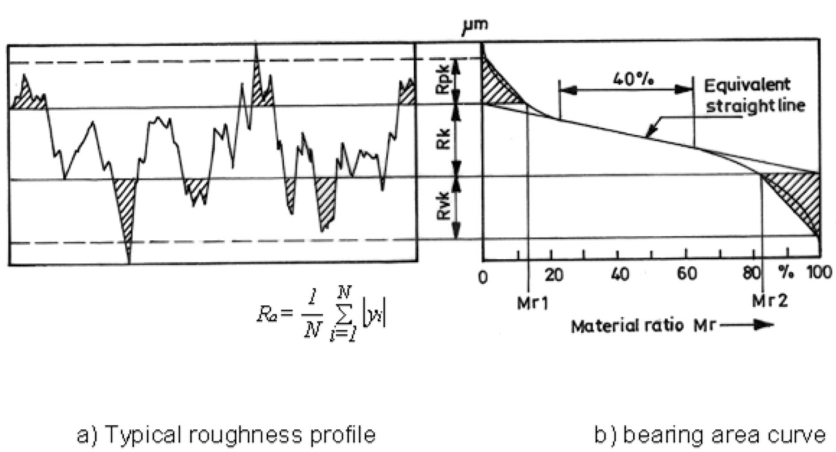


Figure 6. Roughness average (R_a) and bearing area parameters



ensures good sliding properties of smooth surfaces (characterized by R_k) and has greater ability to retain oil in the grooves (characterized by M_{r2}) for effective lubrication. It can be noticed that both R_k and M_{r2} are conflicting in nature. Therefore, achieving required M_{r2} and R_k within range 75-85% and 0.6-1.8 µm respectively as prescribed by the engine manufacturers in the final cylinder liners are important in plateau honing operation. If the honing parameters are not properly selected from the manufacturing point of view, it is very difficult to achieve the bearing area parameters within the specified range in the final components. Therefore, these constraints leaves very difficult for the manufacturing engineers to select the appropriate plateau honing process parameters. Though different techniques are available, there are only a limited attempts made by researchers to select suitable honing parameters considering the surface topography, especially the bearing area parameters. Since the NSGA has emerged as one of the most efficient multi-objective evolutionary algorithm for solving multi-obective problems, the NSGA applied to a plateau honing process has been detailed in the next section.

3.4 Application of NSGA for Plateau Honing Process

The following sections deals with the selection of Pareto-optimal solutions using NSGA for plateau honing process considering it as a multi-objective problem. The input honing parameters considered are rotary speed (RS) reciprocating speed (ST), pressure (PR) and honing time (HT) (Table 2). The output parameters considered are M_{r2} and R_k. The procedure of NSGA applied to plateau honing process is given in Figures 7 and 8 for easier understanding. The results obtained through NSGA of the 200th iteration are given in Table 3.

3.4.1 Selection of Chromosome and Coding

Initially, the chromosomes for the entire population are randomly generated to allow different solutions to be scattered in the search space. A typical chromosome obtained from the selected population is shown in Figure 7a and also in Table 3a (S.No.1). The input honing parameters used in the trials have different lower and upper limits with different step lengths (Table 2). Therefore, a binary coding with '0' and '1' is used to represent the variable honing parameters. The total length of the string is 16, in which first 4 bits represent rotary speed (RS), the second 4 bits represent reciprocating speed (ST), the third 4 bits represent honing pressure (PR) and fourth 4 bits represents honing time (HT) (Figure 7a).

The binary numbers are decoded using the formula

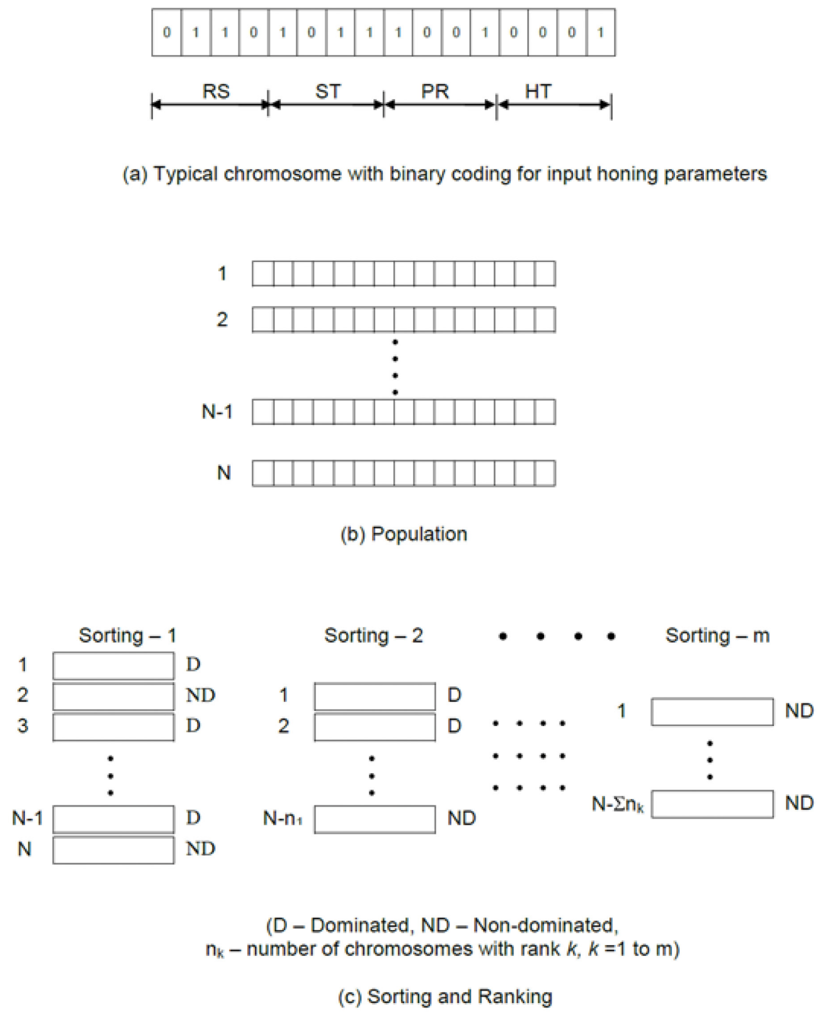
$$X_A = X_L + \frac{X_U - X_L}{2_n - 1}(Decoded\ \ value) \tag{2}$$

where X_A is the actual value of honing parameters, X_U and X_L are the upper and lower limits of the individual honing parameters and *n* is the sub-

Table 2. Honing parameters, limits, step, and bit length used for NSGA

S. No	Honing parameters	Lower limit	Upper limit	Step length	Bit length
1	Rotary speed (RS), m/min.	34	40	0.400	4
2	Rec. speed (ST), m/min.	10.4	14.4	0.266	4
3	Pressure (PR), MPa	0.5	0.7	0.013	4
4	Honing time (HT), s	8	16	0.533	4

Figure 7. The procedure of NSGA (Part 1)



string length. Typical decoded and actual values of the 200th generation are shown in Table 3, Col. (a) and Col.(b) respectively.

3.4.2 Sorting and Ranking

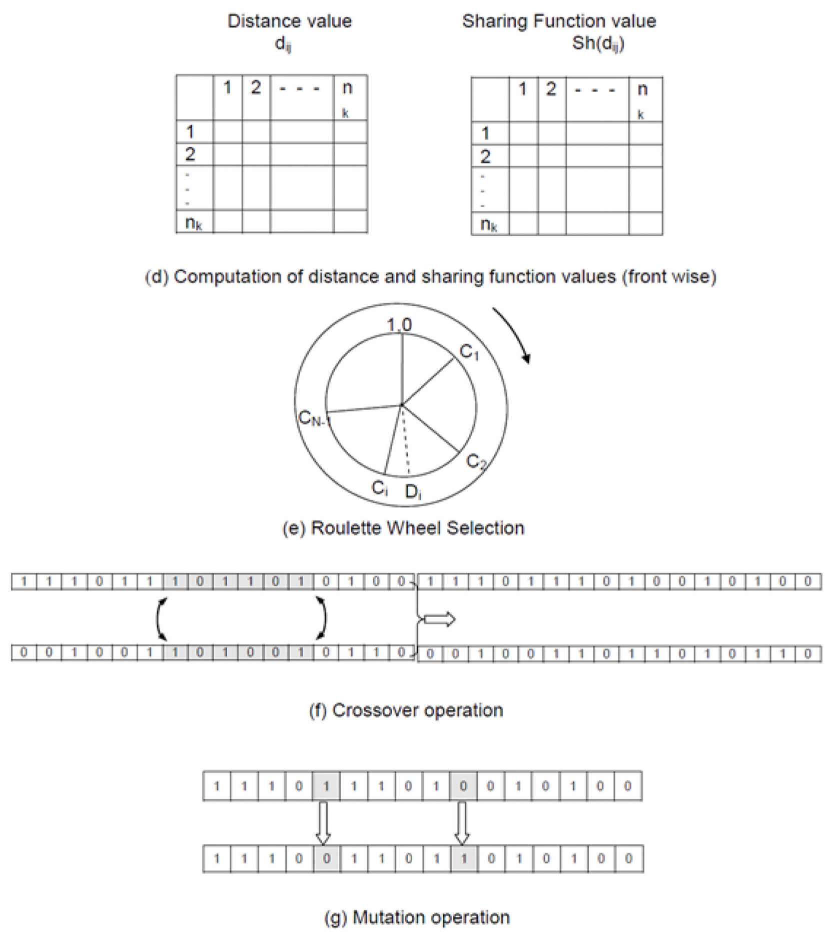
After generating number of populations (N=80) (Figure 7b and Table 3a), the chromosomes are sorted to assign ranks. The ranking is done based on its closeness towards optimizing both the fitness (objective) functions. The fitness functions used in this work are as follows.

$$Fn.1 = 1/|M_{r2(min)} - M_{r2}| + |M_{r2} - M_{r2(max)}| \quad (3)$$

$$Fn.2 = 1/|R_{k(min)} - R_{k}| + |R_{k} - R_{k(max)}| \quad (4)$$

The values of the fitness functions are computed for each string in the population (Table 3 Col.(c)). The M_{r2} and R_k values for the fitness Equations 3 and 4 are obtained from the multiple linear regression Equations 5 and 6 respectively. The multiple regression equations are obtained from the experimental data (Kanthababu et al., 2009a)

Figure 8. The procedure of NSGA (Part 2)



$$M_{r2} = ST^{0.2170}\, PR^{-0.1648}\, HT^{0.0493}\, RS^{1.0083} \quad (5)$$

$$R_k = ST^{0.0303}\, PR^{-0.6805}\, HT^{-1.0651}\, RS^{0.6503} \quad (6)$$

The objective of sorting and ranking is to find a string with maximum fitness function value of Fn. 1 and Fn. 2. The denominator in the Equations 3 and 4 represents transformed parameters. By such transformation, denominator reaches a minimum value when the parameter is within the specified range. The minimum value is equal to the range itself. For example, if the range for M_{r2} (Table 4) is used in the Equation 3, i.e. $M_{r2(min)}$ is 75% and $M_{r2(max)}$ is 85%, then M_{r2} of 80% results in a denominator value of 10 and the therefore the maximum value of Fn.1 is 0.1. If any value of M_{r2} less than 75% or greater than 85% gives a denominator value greater than 10 and as a result, the maximum value of Fn.1 will be less than 0.1. Similarly, if the range for R_k (Table 4) is used in the Equation 4 i.e. $R_{k(min)}$ is 0.8 and $R_{k(max)}$ is 1.6 µm, then it results in maximum value of 1.25 for Fn.2 (Table 4). The maximum and minimum values for M_{r2} and R_k have been fixed based on the recommendations of the engine manufacturers for effective performance of the engine (Table 4).

In order to maximize both the fitness functions, the ranking is done based on Equation 7 as given.

Table 3. Results of NSGA of 200th iteration (only first 20 are given)

S. No.	Chromosome (a)	Honing parameter (b)				Fitness function (c)		Rank (d)	Output parameter (e)	
		RS (m/min.)	ST (m/min.)	PR (MPa)	HT (s)					
						Fn. 1	Fn. 2		Mr_2 (%)	R_k (μm)
1	0110101110010001	36.40	13.33	0.62	8.53	0.57	1.31	3	79.13	1.58
2	0011111100100111	35.20	14.40	0.53	11.73	0.43	10.00	1	81.16	1.23
3	1010000010010000	38.00	10.40	0.62	8.00	0.26	0.95	6	78.05	1.73
4	0010011010010110	34.80	12.00	0.62	11.20	0.10	9.22	2	74.91	1.15
5	1010111100110000	38.00	14.40	0.54	8.00	0.09	0.70	13	85.68	1.92
6	0000000101010000	34.00	10.67	0.57	8.00	0.06	0.98	11	71.20	1.71
7	0100110100011000	35.60	13.87	0.51	12.27	0.26	10.00	1	81.95	1.21
8	0010111001110001	34.80	14.13	0.59	8.53	0.17	1.30	7	77.14	1.58
9	1001011000101010	37.60	12.00	0.53	13.33	0.13	6.09	5	83.91	1.12
10	1010000101110100	38.00	10.67	0.59	10.13	1.00	2.69	1	79.97	1.39
11	0011011011001110	35.20	12.00	0.66	15.47	0.13	1.20	9	76.20	0.78
12	0110111100010001	36.40	14.40	0.51	8.53	0.17	0.83	10	82.99	1.80
13	0111011011101111	36.80	12.00	0.69	16.00	0.73	1.13	3	79.31	0.76
14	0101011111001111	36.00	12.27	0.66	16.00	0.32	1.16	4	78.45	0.77
15	0111011110010100	36.80	12.27	0.62	10.13	0.66	4.08	1	79.24	1.32
16	0111001100001011	36.80	11.20	0.50	13.87	0.29	4.67	3	81.74	1.09
17	1110011001110000	39.60	12.00	0.59	8.00	0.11	0.78	12	84.54	1.84
18	0010100101110010	34.80	12.80	0.59	9.07	0.12	1.78	9	75.72	1.48
19	0110001011001011	36.40	10.93	0.66	13.87	0.16	1.66	7	76.83	0.90
20	1011011101111000	38.40	12.27	0.59	12.27	0.12	8.78	3	84.10	1.14

continued on following page

Table 3. Continued

S. No.	Niche count (f)	Assigned fitness (g)	Shared fitness (h)	Expected count (i)	Probability (j)	Cumulative probability (k)	Random number (l)	Selected S. No. (m)
1	1.27	20.50	16.12	1.317	0.016	0.016	0.813	62
2	1.90	80.00	42.06	3.436	0.043	0.059	0.082	4
3	1.00	2.62	2.62	0.214	0.003	0.062	0.734	60
4	1.18	24.47	20.70	1.691	0.021	0.083	0.210	15
5	1.00	0.70	0.70	0.057	0.001	0.083	0.608	46
6	1.00	0.30	0.30	0.024	0.000	0.082	0.636	48
7	1.88	80.00	42.65	3.484	0.044	0.126	0.691	56
8	1.51	1.22	0.81	0.066	0.001	0.127	0.394	27
9	1.00	7.09	7.09	0.579	0.007	0.134	0.780	62
10	2.49	80.00	32.12	2.624	0.033	0.167	0.910	69
11	1.96	0.20	0.10	0.008	0.000	0.167	0.388	27
12	1.00	0.10	0.10	0.008	0.000	0.167	0.542	44
13	1.00	20.50	20.50	1.674	0.021	0.188	0.145	10
14	1.00	7.29	7.29	0.595	0.007	0.195	0.055	2
15	1.80	80.00	44.53	3.638	0.045	0.240	0.496	38
16	1.20	20.50	17.06	1.394	0.017	0.258	0.667	55
17	1.00	0.50	0.50	0.041	0.001	0.257	0.583	46
18	1.00	0.20	0.20	0.016	0.000	0.258	0.266	20
19	1.25	1.22	0.97	0.079	0.001	0.259	0.447	32
20	1.00	20.50	20.50	1.674	0.021	0.280	0.551	46

continued on following page

Table 3. Continued

S. No.	Mating chromosome (n)	Crossover (yes-1/no-0) (o)	Crossover site 1 (p)	Crossover site 2 (q)	After crossover (r)	Mutation site(s) (s)	After mutation (t)
1	1110111011010100	1	7	12	1110111010010100	5, 10	1110011011010100
2	0010011010010110	1	7	12	0010011011010110	7, 13	0010010011011110
3	0011111000010101	1	11	15	0011111000010100	14	1011111000010000
4	0111011110010100	1	11	15	0111011110010101	10	0111011111010101
5	0110011110011011	1	5	8	0110011000011011	5, 12, 14	0110111000001111
6	0110011000010111	1	5	8	0110011110010111	6, 16	0110001110010110
7	0100111010010101	0	0	0	0100111010010101	8, 14	0100111110010001
8	0101011001101101	0	0	0	0101011001101101	3, 13,15	0111011001100111
9	1110111011010100	1	4	13	1110001100011000	-	1110001100011000
10	0111001100011001	1	4	13	0111111011010101	12,16	0111111011000100
11	0101011001101101	0	0	0	0101011001101101	10,12	0101011000111101
12	1111001011010100	0	0	0	1111001011010100	15	1111001011010110
13	1010000101110100	1	4	7	1010111101110100	-	1010111101110100
14	0011111100100111	1	4	7	0011000100100111	7	0011001100100111
15	1010001110010100	1	4	16	1010110110001000	11	1010110110101000
16	0110110110001001	1	4	16	0110001110010100	2	0010001110010100
17	0110011110011011	1	13	14	0110011110011001	2, 11, 15	0010011110111011
18	1011011101111000	1	13	14	1011011101111010	-	1011011101111010
19	1110001011100110	1	8	16	1110001010011010	3	1100001010011010
20	0110011110011011	1	8	16	0110011111100110	-	0110011111100110

Table 4. Fitness function values for M_{r2} and R_k

Output	Min. value	Max. value
M_{r2} (%)	75	85
R_k (µm)	0.8	1.6
Max. Fn. 1	0.1	
Max. Fn. 2	1.25	

$$\text{Fn.1}[i] < \text{Fn.1}[j] \text{ and } \text{Fn.2}[i] < \text{Fn.2}[j],\ i \neq j \quad (7)$$

where i and j are chromosome numbers. The procedure for sorting and ranking are detailed below.

In the first sorting, each set is checked whether it satisfies the rules as per Equation 7 with respect to the other set in the population. If the above rules are satisfied for any one of the remaining values, then the set is marked as dominated, otherwise it is marked as non-dominated (Figure 7c). All the non-dominated sets obtained in the first sorting are ranked as 1. The remaining (dominated) sets are again sorted out and rank 2 set is obtained. This procedure is repeated until all sets are ranked (Table 3, Col.(d)). The non-dominated solutions of rank 1 are better than non-dominated solutions of rank 2 and so on. The input parameters corresponding to rank 1 are the best in terms of their closeness to the Pareto-optimal solutions in the population.

3.4.3 Shared Fitness

In order to ensure that less crowded regions in the front are adequately emphasized, sharing fitness method is used. For calculating the shared fitness, the distances and a sharing function of each chromosome with respect to the other chromosomes are calculated front wise. Then, a niche value for each chromosome is calculated based on the sharing function values. The calculation of shared fitness is as follows: The population in the first front set (rank 1) is initially assigned a temporary (dummy) fitness value of 80 (equal to population size) (Table 3, Col.(g)). Then the normalized Euclidean distance (d_{ij}) (Equation 8) of each chromosome is calculated with respect to all other chromosomes within the first front set using the formula

$$d_{ij} = \sqrt{\sum_{s=1}^{p_1} \left(\frac{x_s^{(i)} - x_s^{(j)}}{x_s^{\max} - x_s^{\min}} \right)^2} \quad (8)$$

where p_1 is total number of input parameters, x_s is the value of s^{th} input parameter with upper and lower bound values, while *i* and *j* are chromosome numbers (Figure 7d). Then the sharing function values ($Sh(d_{ij})$) of all the first front chromosomes are calculated using the formula

$$Sh(d_{ij}) = \begin{cases} 1 - \left(\dfrac{d_{ij}}{\sigma_{share}} \right)^2, & if\ d_{ij} < \sigma_{share} \\ 0, & otherwise \end{cases} \quad (9)$$

where σ_{share} is the maximum distance allowed between any two chromosomes to become members of a niche (Figure 7d). The σ_{share} value in the Equation 9 is chosen as 0.5. The sharing function takes a value between zero and one depending on the distance d_{ij}. Then niche count (nc_i) (Table 3, Col.(f)) provides an estimate of the extent of crowding near a chromosome and is calculated using the Equation 10.

$$nc_i = \sum_{j=1}^{N} Sh(d_{ij}) \quad (10)$$

where *N* is the total population. If there is no solution within a radius of σ_{share} then the niche count for that solution is taken as one. Then the shared fitness values (F_i') (Table 3, Col.(h)) are calculated by dividing the temporary fitness values by the niche count, which is given by

$$F_i' = \frac{F_i}{nc_i} \quad (11)$$

After calculating the shared fitness values for the first front, a smaller value of 0.2 is subtracted from the minimum shared fitness value of the population in the first front and the resultant value is assigned as the temporary fitness value for all the second front (rank 2) chromosomes. The above procedure is adopted for calculating also the shared fitness value for the second front. This has been done to make sure that no solution in the first front has been assigned with fitness value worse than the second front. This procedure is continued to find the shared fitness values for all the fronts (Table 3, Col.(h)).

3.4.4 Roulette Wheel Selection

After calculating the shared fitness values for all the chromosomes, a new mating pool (Table 5, Col.(n)) is selected from the given population based on the shared fitness values using roulette wheel selection operator (Figure 7e). The roulette wheel selection procedure is as follows:

First, the expected count A_i (Table 3, Col.(i)) is calculated by

$$A_i = \frac{F_i'}{\bar{F}_i'} \quad (12)$$

$\bar{F}_i'$ is the average fitness value given by

$$\bar{F}_i' = \frac{\sum F_i'}{N} \quad (13)$$

where N is the total population. Then the probability of selection (Table 3, Col.(j)) and cumulative probability (Table 3, Col.(k)) are calculated by

$$B_i = \frac{A_i}{N} \text{ and } C_i = \sum B_i \quad (14)$$

The cumulative probability will add up to 1 for the N^{th} (80^{th}) chromosome. To find mating chromosome, random numbers between 0 and 1 are generated N times (Table 3, Col.(l)). Each random number is checked for the range of cumulative probability within which it is lying. Then the chromosome corresponding to the next

Table 5. Results of Pareto-optimal set

S. No.	Input honing parameters				Fitness function		Predicted output		Rank
	RS (m/min.)	ST (m/min.)	PR (MPa)	HT (s)	Fn. 1	Fn. 2	M_{r2} (%)	R_a (μm)	
2	36.8	13.87	0.55	15.47	0.10	1.25	84.66	0.91	1
12	39.6	12.80	0.57	10.13	0.07	1.25	87.39	1.48	1
23	39.6	13.60	0.55	12.27	0.05	1.25	89.74	1.23	1
35	37.2	11.20	0.57	8.53	0.10	1.01	79.04	1.70	1
40	38.8	10.40	0.55	8.53	0.10	0.88	81.47	1.77	1
52	39.6	11.20	0.54	13.33	0.07	1.25	86.74	1.13	1
58	38.0	11.73	0.69	11.20	0.10	1.25	80.10	1.13	1
69	34.4	13.87	0.67	8.00	0.09	1.25	74.13	1.55	1
76	36.0	13.07	0.54	8.00	0.10	0.77	79.44	1.85	1
79	38.0	11.20	0.54	8.00	0.10	0.71	81.14	1.90	1

upper bound of the cumulative probability range is selected for the mating pool (Table 3, Col.(m)). Thus a population of N^{th} chromosomes is obtained for the mating pool (Table 3, Col.(n)). Therefore, each solution in the first rank has a better chance of surviving in the mating pool than that of the second front and so on.

3.4.5 Crossover and Mutation

After selecting chromosomes for the mating pool, a randomly generated two-point crossover (Table 3, Col.(p) and Col.(q)) is adopted for crossing with a crossover probability of 0.8 (Figure 7f). The crossover probability has usually a higher value, as generation of new chromosomes is to be encouraged. Then bitwise mutation is performed subsequently with a mutation probability of 0.1, which is commonly taken to be a smaller value (Table 3, Col.(s) and Figure 7g). The chromosomes obtained after mutation in the first generations are used as input for next generations and so on (Table 3, Col.(t)).

The above procedure is continued until the termination criterion (200 generation) is over. The values shown in Table 3 are obtained after the 200th generation. The finally generated chromosome is expected to have better solutions than that of the previous generations (Table 3, Col.(t)). The chromosomes obtained in the final generation with rank 1 are considered as Pareto-optimal solutions. Typical results of the Pareto-optimal solutions (1st rank) obtained at the 200th iteration is shown in Table 5. The NSGA algorithm is implemented using TurboC and run on Pentium IV PC.

The Pareto-optimal set or non-dominated solutions (Table 5) obtained through NSGA will serve as a ready reference for the manufacturer of cylinder liners to implement in plateau honing operation in order to achieve the multiple objectives M_{r2} and R_k values within certain limits. The same procedure could be used for any manufacturing processes by clearly defining the objective functions.

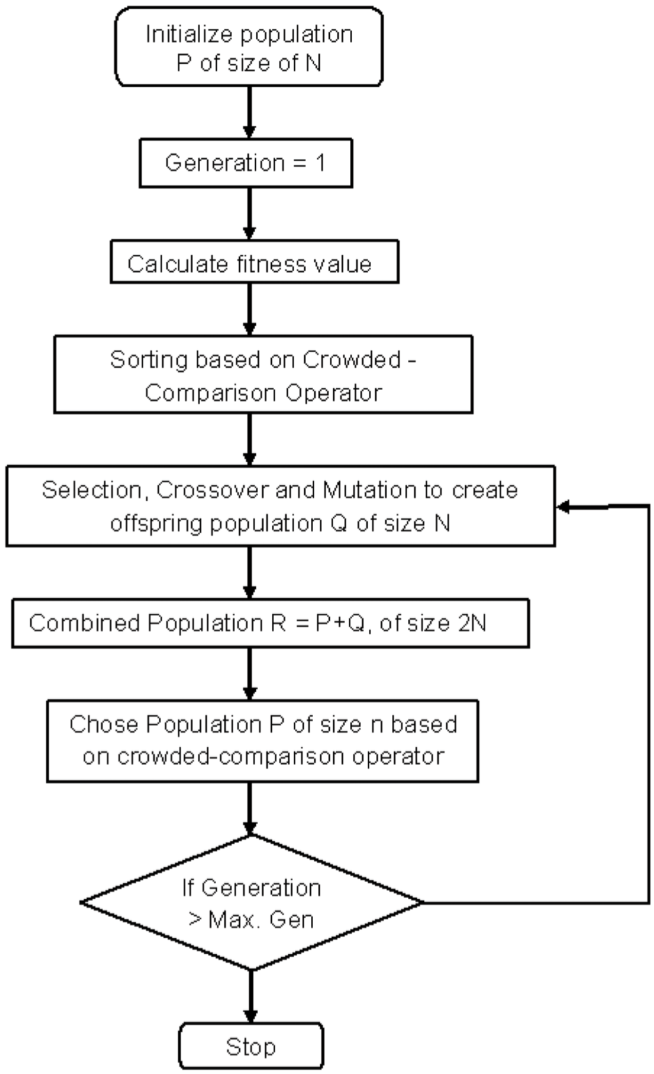
3.5 Non-Dominated Sorting Genetic Algorithm-II (NSGA-II)

NSGA-II is proposed by Deb et al. (2002). It is fast and further advancement of simple NSGA and addresses certain issues like high computational complexity of non-dominated sorting, lack of elitism, need for specifying the sharing parameter, etc. The flowchart of NSGA-II is given in Figure 9. In NSGA-II, the crossover and mutation operators remain same but the selection operator works differently from simple GA. In this algorithm, selection is carried out with the help of crowded-comparison operator based on the ranking and crowding distance. The offspring population is first created by using the parent population. Instead of finding the non-dominated front, the two populations are combined to form of size 2N. Then, a non-dominated sorting is used to classify the entire population. The main advantage of maintaining non-dominated solutions in the population is straightforward implementation. In this strategy, the population size is an important GA parameter, since no external archive is used. Although this requires more effort, however it allows a global non-dominance check among the parent and offspring solutions. Once the non-dominated sorting is over, the new population is filled by solutions of different non-dominated front, one at a time, starting with the best non-dominated front, followed by the next best non-dominated front and so on. Since the population size is 2N, not all fronts may be accommodated in the new population. All fronts which could not be accommodated are deleted. The detailed procedure of NSGA-II is given in Deb et al. (2002) and Konak et al. (2006).

4. FUTURE DIRECTIONS OF EVOLUTIONARY ALGORITHMS

Many engineering problems have multiple objectives including engineering system design, reli-

Figure 9. Flowchart of NSGA-II (Mandal et al., 2007)



ability optimization, telecommunication networks, etc. Therefore, evolutionary algorithms created interest among researchers and manufacturing engineers to solve multi-objective problems in real world applications. As the interests grow, there is a need for a focused research and application. The following are some of the immediate research directions that may help to develop better evolutionary algorithms for multi-objective optimization problems.

1. Comparison of existing multi-objective GAs
2. Understand the dynamics of GA populations
3. Multi-objective GAs with number of objectives
4. Convergence to Pareto-optimal front
5. Hybrid multi-objective GAs
6. Real-world application problems

5. CONCLUSION

In this chapter, the importances of evolutionary algorithms which have been evolved as useful tool to solve real world multi-objective problems have been discussed. By evolving a population of solutions, evolutionary algorithms are able to approximate the Pareto-optimal set in a single run without restricting to local optimal solution. Different evolutionary algorithms which have been developed in order to optimize conflicting multiple objectives by researchers for a variety of applications have been detailed by focusing on their components and their salient issues encountered while implementing. Among the multi-objectives evolutionary algorithms, the NSGA and NSGA-II have been found to possess many advantages and successfully applied to optimize different manufacturing processes with conflicting objectives. The concepts and working mechanisms of NSGA have been elucidated with a case study for optimizing multiple objectives in plateau honing process. The future scope for the research and the areas in which it can be applied have been discussed.

ACKNOWLEDGMENT

The author would like to acknowledge the services of Mr. R. Venkataraman, Mr. M. Prakash, Mr. A. Gopikrishna, Mr. D. Sathish Kumar, Department of Manufacturing Engineering, Anna University, Chennai, India for the preparation of this chapter and special thanks to my dear friend Dr. S. Arivazhagan for his constant source of encouragements, valuable guidance and suggestions.

REFERENCES

Davis, J. R. (1989). *Machining: Honing. ASM International Handbook Committee*. USA: Metals Park.

Deb, K. (1995). *Optimization for engineering design: Algorithms and examples*. New Delhi, India: Prentice-Hall.

Deb, K. (1999). Evolutionary algorithms for multi-criterion optimization in engineering design. In Miettinen, K., Mäkelä, M., Neittaanmäki, P., & Périaux, J. (Eds.), *Proceedings of Evolutionary Algorithms in Engineering and Computer Science (EUROGEN-99)* (pp. 135–161). Jyväskylä, Finland.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Singapore: John Wiley & Sons.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2002). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*(2), 182–197. doi:10.1109/4235.996017

Feng, C.-X., Wang, X., & Yu, Z. (2003). Neural network modeling of honing surface roughness parameters defined by ISO 13565. *Journal of Manufacturing Systems*, *21*(5), 395–408. doi:10.1016/S0278-6125(02)80037-1

Fonseca, C. M., & Fleming, P. J. (1993, July,). *Genetic algorithms for multi-objective optimization: Formulation, discussion and generalization*. Paper presented at the 5th International Conference on Genetic Algorithms (pp. 416-423). San Mateao, CA. Morgan Kaufmann.

Fonseca, C. M., & Fleming, P. J. (1995). An overview of evolutionary algorithms in multi-objective optimisation. *Evolutionary Computation Journal*, *3*(1), 1–6. doi:10.1162/evco.1995.3.1.1

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley.

ISO13565-2:1996. (1996). *Geometrical product specifications (GPS) - Surface texture: Profile method- Surfaces having stratified functional properties - Part 2: Height characterization using the linear material ratio curve.* Switzerland: ISO.

ISO 11562:1996. (1996). *Geometrical product specifications (GPS) - Surface texture: Profile method - Metrological characteristics of phase correct filters*. Switzerland: ISO.

ISO 13565-1:1996. (1996). *Geometrical product specifications (GPS) - Surface texture: Profile method- Surfaces having stratified functional properties. Part 1: Filtering and general measurement conditions*. Switzerland: ISO.

Kanthababu, M., Shunmugam, M. S., & Singaperumal, M. (2009a, May). *Multi-objective optimization of plateau honing process parameters using non-dominated sorting genetic algorithm*. Paper presented at the 12th CIRP International Conference on Modeling of Machining Operations (pp. 605-612). Donostia-San Sebastián, Spain.

Kanthababu, M., Shunmugam, M. S., & Singaperumal, M. (2009b). Identification of significant parameters and appropriate levels in honing of cylinder liners. *International Journal of Machining and Machinability of Materials*, *5*(1), 80–96. doi:10.1504/IJMMM.2009.023114

Kanthababu, M., Shunmugam, M. S., & Singaperumal, M. (2010). Data mining approach for selection of plateau honing parameters with dual objectives. *International Journal of Machining and Machinability of Materials*, *8*(1/2), 167–176. doi:10.1504/IJMMM.2010.034494

Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, *91*, 992–1007. doi:10.1016/j.ress.2005.11.018

Kuriakose, S., & Shunmugam, M. S. (2005). Multi-objective optimization of wire-electro discharge machining process by nondominated sorting genetic algorithm. *Journal of Materials Processing Technology*, *170*, 133–141. doi:10.1016/j.jmatprotec.2005.04.105

Mandal, D., Pal, S. K., & Saha, P. (2007). Modeling of electrical discharge machining process using back propagation neural network and multi-objective optimization using non-dominating sorting genetic algorithm-II. *Journal of Materials Processing Technology*, *186*(1-3), 154–162. doi:10.1016/j.jmatprotec.2006.12.030

Nandasana, A., Ray, D., Kumar, A., & Santosh, K. G. (2003). Applications of the non-dominated sorting genetic algorithm (NSGA) in chemical reaction engineering. *International Journal of Chemical Reactor Engineering*, *1*, R2. doi:10.2202/1542-6580.1018

Pawlus, P. (1994). A study on the functional properties of honed cylinder surface during running-in. *Wear*, *176*, 247–254. doi:10.1016/0043-1648(94)90153-8

Pawlus, P. (1997). Change of cylinder surface topography in the initial stage of engine life. *Wear*, *209*, 69–83. doi:10.1016/S0043-1648(97)00007-0

Pawlus, P. (1998). Digital filtering for surface texture measurement on inner cylindrical surfaces. *Measurement*, *24*, 139–159. doi:10.1016/S0263-2241(98)00050-5

Rao, S. S. (1984). *Optimization theory and application*. New Delhi, India: Wiley Eastern Limited.

Sardiňas, R. Q., Reis, P., & Davim, J. P. (2006). Multi-objective optimization of cutting parameters for drilling laminate composite materials by using genetic algorithms. *Composites Science and Technology*, *66*(15), 3083–3088. doi:10.1016/j.compscitech.2006.05.003

Sbalzariniy, I. F., Muller, S., & Koumoutsakos, P. (2000). Multi-objective optimization using evolutionary algorithms. *Proceedings of the Center for Turbulence Research Summer Program*, *2000*, 63–74.

Schaffer, J. D. (1985). Multiple objective optimization with vector evaluated genetic algorithms. *1st IEEE International Conference on Genetic Algorithms and Their Applications* (pp. 93-100). Pittsburgh, PA.

Srinivas, N., & Deb, K. (1995). Multi-objective function optimization using non-dominated sorting genetic algorithms. *Evolutionary Computation*, *2*(3), 221–248. doi:10.1162/evco.1994.2.3.221

Valenzuela, C. L. (2002, May). A simple evolutionary algorithm for multi-objective optimization (SEAMO). *IEEE World Congress on Evolutionary Computation (CEC)* (pp. 717-722). Honolulu, Hawaii.

Wang, Z. G., Wong, Y. S., Rahman, M., & Sun, J. (2006). Multi-objective optimization of high-speed milling with parallel genetic simulated annealing. *International Journal of Advanced Manufacturing Technology*, *31*(3-4), 209–218. doi:10.1007/s00170-005-0191-x

Yang, S. H., & Natarajan, U. (2010). Multi-objective optimization of cutting parameters in turning process using differential evolution and non-dominated sorting genetic algorithm-II approaches. *International Journal of Advanced Manufacturing Technology*, *49*(5-8), 773–784. doi:10.1007/s00170-009-2404-1

Zhoua, A., Qu, B.-Y., Li, H., Zhaob, S.-Z., Suganthan, P. N., & Zhangd, Q. (2011). *Multiobjective evolutionary algorithms: A survey of the state of the art*. Swarm and Evolutionary Computation.

Zitzler, E., Deb, K., & Thiel, L. (2000). Comparison of multi-objective evolutionary algorithms: Empirical results. *Evolutionary Computation*, *8*(2), 173–195. doi:10.1162/106365600568202

Zitzler, E., Laumanns, M., & Bleuler, S. (2004). A tutorial on evolutionary multiobjective optimization. In Gandibleux, X., Servanx, M., Sorensen, K., & Kindt, V. T. (Eds.), *Metaheuristics for Multiobjective Optimisation, Lecture Notes in Economics and Mathematical Systems* (pp. 3–37). Berlin, Germany: Springer. doi:10.1007/978-3-642-17144-4_1

KEY TERMS AND DEFINITIONS

Chromosome: In genetic algorithms, a chromosome (also sometimes called a genome) is a set of parameters which define a proposed solution to the problem that the genetic algorithm is trying to solve.

Crossover: In genetic algorithms of computing, crossover is a genetic operator used to vary the programming of chromosome from one generation to the next. It is analogous to reproduction and biological crossover, upon which genetic algorithms are based.

Evolutionary Algorithm: It is a subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm. An Evolutionary algorithm uses mechanisms inspired by biological evolution such as reproduction, mutation, recombination and selection.

Fitness Function: The objective function where the user wants to minimize or maximize. For standard optimization algorithms this is known as the objective function.

Genetic Algorithm: It is a method for solving both constrained and unconstrained optimization problems that is based on natural selection.

Multi-objective Optimization: It is the process of simultaneously optimizing two or more conflicting objectives subject to certain constraints.

Mutation: In genetic algorithms of computing, mutation is a genetic operator used to maintain genetic diversity from one generation of a population of algorithm chromosomes to the next. It is analogous to biological mutation.

Pareto-optimal Solution: Pareto-optimal solutions are optimal compromises of the two criteria in the sense that any improvement of one criterion implies an impairment to the other.

Plateau Honing: It is super-finishing process that generates a surface texture that has the benefits of both smooth surface and rough surface (grooves with cross hatch lay).

Roulette-wheel Selection: It is also known as fitness proportionate selection. It is a genetic operator used in genetic algorithms for selecting potentially useful solutions for recombination.

Surface Texture: It is the characteristics of a surface and also known as surface finish. It has three components namely lay, surface roughness, and waviness.

This work was previously published in Computational Methods for Optimizing Manufacturing Technology: Models and Techniques, edited by J. Paulo Davim, pp. 44-66, copyright 2012 by Engineering Science Reference (an imprint of IGI Global).

Chapter 22
Decision Support Framework for the Selection of a Layout Type

Jannes Slomp
University of Groningen, The Netherlands

Jos A.C. Bokhorst
University of Groningen, The Netherlands

ABSTRACT

One of the most important design decisions in a firm is the choice for a manufacturing layout type. This chapter shows which aspects have to be taken into account and suggests a systematic method for the decision problem. The method can be seen as a decision support framework, which links the various aspects. The framework is based on the AHP (Analytic Hierarchy Process) approach. A case study, concerning a Dutch firm, illustrates the applicability of the framework in a practical instance.

INTRODUCTION

The choice for a manufacturing layout is a strategic issue and has a significant impact on the performance of the operations function of a company (Meijers and Stephens, 2004, Francis et al. 1992). A variety of manufacturing layout types may be applicable in a practical situation. Table 1 presents some alternative layout types for high-variety/low-volume situations. The most dominant layout type in practice is the process-oriented functional layout, where machines of the same type are located in the same area (Slomp et al., 1995). An important alternative is the socalled Cellular Layout type, where machines are grouped in cells and each cell is responsible for the complete manufacturing of a part family. This product-oriented layout type has gained substantial attention in literature and in practice (Wemmerlöv and Hyer, 1989, and Wemmerlöv and Johnson, 1997). Both types of manufacturing layout have their advantages and disadvantages. Several authors present alternative layout types to cope with the disadvantages of the functional and/or cellular layout type. Rosenblatt (1986) suggested a dynamic plant layout where cellular configurations periodically change depending on the demand in each period. Balankrishnan and

DOI: 10.4018/978-1-4666-1945-6.ch022

Cheng (1998) present a review on the dynamic plant layout problem. Venkatadri et al. (1997) and Montreuil et al. (1999) propose a socalled fractal layout for job shop environments in order to gain the flow time advantages of Cellular Manufacturing and the flexibility of a functional layout. This type of layout is robust with respect to changes in demand and product mix. Another robust design, the socalled holographic or holonic layout, is proposed by Montreuil et al. (1993). Here individual machines, or machines types, are strategically distributed through the facility. Production orders are assigned to available machines which are located in the same area of the plant. A special case of the holonic layout is the socalled distributed layout (Benjaafar and Sheikhzdeh, 2000 and Benjaafar et al., 2002) where machine replicates are strategically distributed across physical space. Some researchers stress the need for a hybrid layout system which combines several layout types (e.g. Irani, 1993). Irani and Huang (2000) and Benjaafar et al. (2002) define a modular layout in which products have to be manufactured by one or more modules. Each module may have its own internal layout. A modular layout is an example of a hybrid layout. Wemmerlöv and Hyer (1989) show that many companies apply a hybrid layout.

Table 1. Layout types and some major advantages and disadvantages

Type of Layout	Explanation	Major advantages	Major disadvantages
Process Layout or Functional Layout	Machines of the same type are located in the same area.	Routing Flexibility. Specialization in process type.	Complexity of coordination between departments.
Cellular Layout	Machines are grouped in cells and each cell is responsible for the complete manufacturing of a part family.	Short setup times because of the dedication of families to cells.	Sensitive for unbalance in the load of identical machines in different cells. Inflexible for the introduction of new products.
Dynamic Cellular Layout	A reconfigurable cellular layout.	Enables the cell layout to respond to product changes.	Costs of reallocating machines in case of product changes
Fractal Layout	Machines are grouped in various fractals, which are (more or less) identical cells able to produce all products.	Enables the cell layout to deal with changes in product mix.	Limited specialization of workers and machines.
Holonic Layout or Holographic Layout	Each machine (type) is an autonomous entity (holons) and is seemingly random (=random or based upon transition probabilities) located throughout the plant.	Provides efficient process routes for any production order. As orders arrive, routings are constructed by searching for compatibility between order requirements and machine availability, location, and capability.	Complexity of coordination between machine requirements of the various production orders
Distributed Layout or Scattered Layout. (Distributed or scattered layouts can be seen as special cases of the holonic layout)	Distributed or scattered layouts are those where machine replicates are strategically distributed across physical space.	Flexibility of assigning manufacturing orders to available machines which are located in the same area.	Limited specialization of workers and machines. Complexity of coordination.
Hybrid layout	Several layout types exist within one department	Fit between the various characteristics of the product types of a company and the various layout options.	Complexity of planning and control
Modular layout (Modular layouts can be seen as a special case of a hybrid layout)	Machines are clustered in modules. Each module has its own layout and is responsible for a number of operations to be performed on a product	Recognizes the layout needs of the various operations needed per product.	Complexity of the linkage of the various modules

This chapter presents a general decision support framework for the selection of a manufacturing layout type. Our focus lies on the selection of objectives, aspects and contributing elements for the selection problem. The framework applies the AHP (Analytic Hierarchy Process) approach (Saaty, 1980). This approach is useful for multi-criteria decisions where intuitive, qualitative and quantitative aspects play a role. The approach includes a hierarchical decomposition of the decision problem and a further decomposition of each decision level into pairwise comparisons of decision elements. Next, the "eigenvalue" method is used to estimate the relative weights of the decision elements. For a further explanation of the AHP method, we refer to Saaty (1980) or Zahedi (1986). As will be made clear in the remainder of this chapter, the AHP method offers several advantages in the layout type selection problem.

The next section will provide some further background to the selection problem. We will make clear that it is important to approach the layout type selection problem from a strategic viewpoint. In a subsequent section, we will discuss how layout types influence generic objectives of a company. We will then show how various aspects of layouts have an impact on manufacturing performance. The performance objectives and the various aspects are presented in the form of a decision hierarchy, according the AHP approach. After specifying the AHP approach, we present a case study to indicate the generic value of the defined decision hierarchy. The last section of this chapter is meant to reflect on the proposed selection methodology and to draw conclusions.

BACKGROUND

Literature on layout design problems falls into two major categories, algorithmic and procedural approaches (Yang and Kuo, 2003). Algorithmic approaches make use of simplified design constraints and objectives and can be used to generate layout alternatives efficiently (Meller and Gau, 1996). Algorithmic approaches are useful as a step in the design of a detailed layout. They assume the choice of a layout type. Procedural approaches may incorporate the choice of layout type and take care of both qualitative and quantitative objectives in the whole design process (Muther, 1973). A major disadvantage of a procedural approach is its dependence on the subjective judgement of one ore more experienced designers. Furthermore, procedural approaches divide the design problem in several steps which may lead to suboptimality. In order to overcome this suboptimality, designers may develop alternative layouts, based upon different layout types, and a well-working methodology may support the selection of the best layout. This chapter is devoted to the presentation of such a methodology. The methodology recognizes major differences between layout types.

Several authors propose methodologies to simultaneously cope with qualitative and quantitative objectives in the selection of a layout. Cambron and Evans (1991) applied Saaty's Analytic Hierarchy Process (AHP) to consider the problem's multiple objectives. The approach is illustrated by means of a problem involving the layout of a commercial printing and binding facility. Partove and Burton (1992) also propose AHP for layout selection. Yang and Kuo (2002) and Ertay et. al (2006) apply AHP and combine this with the data envelopment analysis (DEA) approach to solve the layout selection problem. Qualitative performance measures were weighted by AHP. DEA was then used to solve the multiple-objective layout problem. Yang and Kuo (2002) used a practical case study, an IC packaging company, to illustrate the efficiency and effectiveness of their methodology. Ertay et. al (2006) illustrate the applicability by means of a case study in which a choice has to be made between 17 alternative layouts for a company producing plastic profiles. Abdi and Labib (2003) apply AHP for the selection of a reconfigurable manufacturing system. Yang and Hung (2007) explore the use of multiple-attribute decision making (MADM) in the selection of an appropriate layout design. Two methods were proposed in solving the case study problem: the technique for order preference by similarity to ideal solution (TOPSIS) and fuzzy TOPSIS. The methodologies for layout selection, as presented in literature, focus on the methods to deal with conflicting quantitative and/or qualitative objectives. Limited attention is devoted to the selection of appropriate objectives and criteria. This chapter pays substantial attention to the choice of objectives and the specification of criteria.

Objectives and criteria in the layout literature are usually just a listing of some relevant elements. Researchers do not link the elements to the various strategic objectives of a company. The evaluation of layout alternatives is based on objectives such as (i) minimizing material handling costs, (ii) improving flexibility for arrangement and operation, (iii) utilising the available area most effectively, and (iv) minimising overall production time (Francis et al, 1992). Raman et. al (2009) distinguish three layout effectiveness factors—facilities layout flexibility (FLF), productive area utilisation (PAU) and closeness gap (CG). They claim that the measurement of these factors enables the decision-maker of a manufacturing enterprise to analyse a layout, based on which they can make decisions towards productivity improvement. They do not link the three factors to generic performance objectives of manufacturing companies. In this chapter we stress the importance of clarifying the link between layout decision criteria and the performance objectives of the company. This is needed to place the layout decision in its strategic context.

A major decision in many companies is the choice between a product-oriented and a process-oriented design philosophy. In a product-oriented design philosophy, machines and workers are grouped according to manufacturing needs of product types. A group of machines and workers is responsible for the complete manufacturing of (a set of) product types. The various groups in a product layout are relatively independent from each other. In a process-oriented design philosophy, machines and workers are grouped according the various functions needed to perform all product types. The functionally based groups are highly dependent on each other. Products flow from group to group. Most studies on the selection of a manufacturing layout implicitly assume a functional, or process-oriented, layout and are concerned with the allocation of the different functional groups. This chapter explicitly recognizes that companies may select a process-oriented or a product-oriented layout type, or a mix of both types. Table 1 gives an overview of possible layout types in a high-variety/low-volume environment. In our viewpoint, the type of layout has an important impact on the various performance objectives of a company.

The choice between a product- and a process-oriented layout is, many times, not an obvious decision. Case studies and survey articles (see e.g. Wemmerlöv and Hyer 1989, Burbidge et al. 1992) illustrate the enormous advantages of the introduction of a product-oriented, cellular manufacturing layout (CML). Other case studies indicate that several firms move from a cellular manufacturing layout towards a process-oriented, functional layout (FL) (see e.g. Slomp 1998, Molleman *et al.* 2002). Numerous simulation studies have been performed in order to compare the performance of a CML and a FL in various situations (for an overview, see Johnson and Wemmerlöv 1996, Argarwal and Sarkis, 1998, or Shambu et al., 1996). These studies indicate important factors, which have to play a role in the layout choice. Johnson and Wemmerlöv (1996), however, state that the simulation studies cannot assist practitioners in making specific choices between existing layouts and alternative cell systems. They indicate various mismatches between the model world and reality and suggest that decisions to change the existing layout should be made on a case-by-case basis for each potential cell application. A general framework, as presented in this chapter, may support managers in the selection of an appropriate layout type. The case study in this paper shows how the framework has supported the managers involved.

THE LINK BETWEEN STRATEGIC OBJECTIVES AND LAYOUT TYPE

An essential condition in the selection of a layout is the ability of decision makers to link the strengths and weaknesses of each layout alternative with the market demands, or performance objectives, with which the firm has to deal. Slack *et al.* (2001) distinguish five major performance objectives: price, quality, speed, flexibility, and dependability. The flexibility objective can be further split in product/service flexibility, mix flexibility and volume/demand flexibility. These objectives have to play an essential role in the selection of a manufacturing layout. We have added the objective "quality of work" to the set of objectives. This criterion is especially important in environments where labor is scarce. In the next subsections, we link the various performance objectives with aspects of the manufacturing layout type. We will refer to the functional (process-oriented) and cellular (product-oriented) layout type, as described in Table 1, in the discussion of the various aspects. Important words in the next subsections are written in italics. These words concern the aspects and elements which need to be dealt with in the layout selection problem. They are summarized in Table 2.

The Impact of Layout Type on the Price of a Product

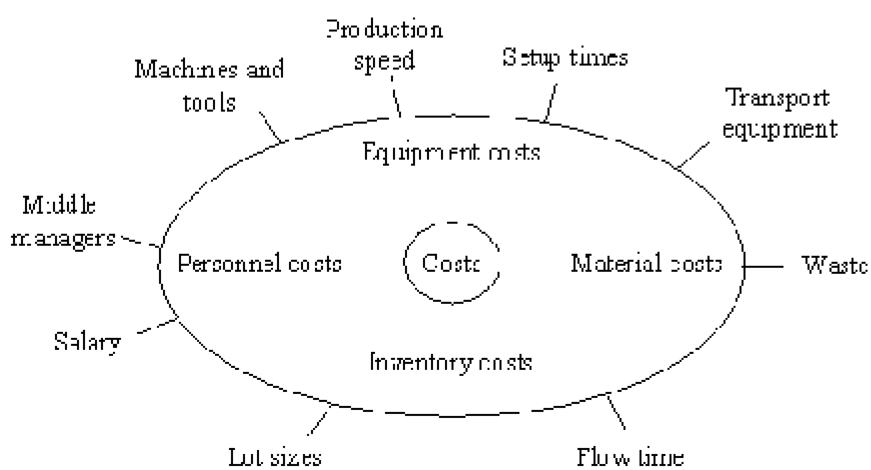
From the perspective of a production manager, the price of a product has to be related to the manufacturing costs. A reduction of the manufacturing costs can be a reason to lower the price of products. Elements of the manufacturing costs are *equipment costs*, *personnel costs*, *material costs* and *inventory costs*. Several aspects of a manufacturing layout will have impact on these

Table 2. Objectives, aspects and elements in the selection of a manufacturing layout

Objectives	aspects	elements
price	costs	• equipment costs • personnel costs • material costs • inventory costs
quality	• specialization of workers • advanced machinery • control loops	
speed	throughput time	• transport time • machining time • waiting time
flexibility • product/service • mix • volume/demand	• response • range	
dependability	• interchangeability of workers • interchangeability of equipment • control capacity	
quality of labor	• skill variety • task identity • task significance • autonomy • feedback	

costs (see Figure 1). More *machines and tools* may be needed in a product-oriented cell layout in order to create independent, autonomous groups of workers and machines. A survey of 32 U.S. firms involved with cellular manufacturing, reported in Wemmerlöv and Hyer (1989), showed that new equipment and machine duplication was a major expense category for cell implementation. Specialization in a process layout may lead to a higher *production speed*, which may reduce equipment and personnel costs. On the other hand, *setup times* are usually lower in a product layout (see e.g. Flynn and Jacobs 1986, Wemmerlöv and Hyer 1989, Wemmerlöv and Johnson 1997) and reduce equipment and personnel costs in this type of layout. Furthermore, less *transport equipment* is usually required in a product layout because of the shorter transport distances. All these aspects have to be considered in order to estimate the impact on the equipment costs by the various types of layout. Personnel costs are, as mentioned above, related to the factors that have impact on equipment costs. Personnel costs in a product layout can also be lower than in a process layout because of a reduced need of *middle managers*. Farrington and Nazemetz (1998) indicate, by means of simulation studies, that a cellular system is easier to manage than a job shop. Empirical studies show the reduced need for indirect labor where firms convert from a functional layout to a cellular layout (Wemmerlöv and Hyer 1989, Burbidge 1992, Slomp *et al.* 1993). On the other hand, the *salaries* in a product-oriented layout may be higher, since more tasks are, probably, decentralized to the autonomous groups and workers need higher qualifications. These aspects of personnel costs should be taken into account when assessing the various layout alternatives. Material costs can be influenced by the layout through the effect of layout choice on *waste*. It is conceivable that workers in a product layout feel more responsibility for the reduction of the amount of waste. On the other hand, more advanced equipment and more specialized workers in a process layout may also reduce waste. Inventory costs can be lower in a product layout because of the smaller *lot sizes* that can be produced efficiently in this type of layout. This efficiency is due to the smaller setup times in a product layout. Further, *flow times* in a product layout are often lower than in a process layout and this reduces the work-in-process inventory. Reductions in throughput time and work-in-process inventory have been reported in surveys of plants that implemented cellular manufacturing (Wemmerlöv and Hyer 1989, Wemmerlöv and Johnson 1997).

Figure 1. Impact of layout factors on manufacturing costs



The Impact of Layout Type on the Quality of a Product

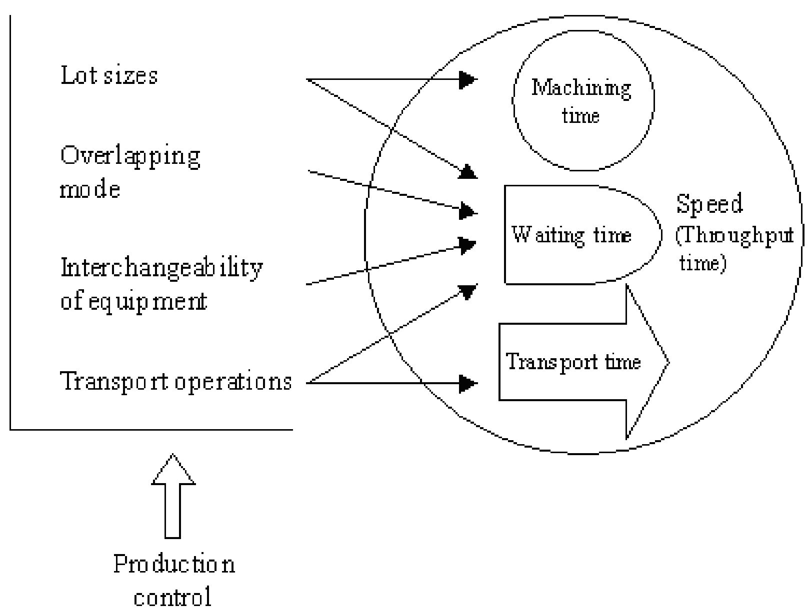
The type of manufacturing layout can also influence the quality of products. In a process layout, workers are probably more *specialized* and will provide for a better product quality. In a product layout, experts are divided among the various groups and the best worker will not always be assigned to the most complex task. Furthermore, a process layout may apply more *advanced machinery*, which has a positive effect on the quality of products. On the other hand, the *control loops* in a product layout are short and may, in comparison to a process layout, have a positive effect on product quality.

The Impact of Layout Type on the Speed to Serve Customers

Speed concerns the time needed to fulfill the needs of internal or external customers. The throughput time of an average job is a measure for speed. This throughput time consists of *transport*, *machining* and *waiting times* (see Figure 2).

A product layout usually involves less *transport operations* for manufacturing jobs because of the proximity of the required machines. It may also be possible to produce in an *overlapping mode* (Shafer and Charnes 1993, Shafer and Meredith 1993), which reduces the "lot size" waiting times. This is more easily realized in a product layout. Problematic for the waiting times in a product layout may be the lack of pooling synergy (Suresh and Meredith 1994). Identical machines are probably split over more than one group and the *interchangeability of equipment* is less than in a process layout. The throughput time will also be influenced by the *lot sizes.* Reduced setup time, which can be realized in a CM environment, may make smaller lot sizes acceptable. This may have a positive impact on the throughput time (Suresh, 1991). *Production control*, finally, plays an important role in the ability to realize short throughput times. An inflexible control system, for instance, may frustrate the production in an overlapping mode. When assessing the effect of

Figure 2. Impact of layout factors on speed



different manufacturing layouts on speed, it is important to consider the requirements with respect to the production control system.

The Impact of Layout Type on the Flexibility to Serve Customer

As mentioned earlier, the flexibility objective can be further split in product/service flexibility, mix flexibility and volume/demand flexibility (Slack *et al.* 2001). The importance of these types of flexibility for a particular instance may differ substantially. Therefore, these types of flexibility have to be seen as different performance objectives. Flexibility, in general, can be defined as the ease (time, effort and/or money) by which changes can be realized. Two aspects determine the flexibility of a manufacturing layout: (i) *range* and (ii) *response*. "Range" refers to the scope of a layout and indicates the variety of situations that can be dealt with without a serious change of the production layout. "Response" indicates the speed by which the layout can be adapted to changing circumstances. Slack (1987) and Upton (1994) have observed that managers think along these lines with respect to the term flexibility.

Product flexibility indicates the ease by which new products can be introduced in a firm. This type of flexibility is higher in a product layout if the new product can be assigned to a single existing product group (quicker response). If a new product has impact on the design of the manufacturing cells, then a process layout is more stable and has more product flexibility (larger range).

Mix flexibility indicates the ease by which a firm can vary the mix of products. Important for the assessment of mix flexibility is insight in the effect of mix changes on the need of various manufacturing processes. A process layout is more range-flexible if the impact of mix changes on the need for manufacturing processes is limited. A product layout is range-flexible if work can be reallocated between the various groups. An important advantage of a product layout concerns the multi-functionality of workers: there are more capabilities at the work floor to deal with changes in the product mix.

Volume/demand flexibility concerns the ease by which the production volume can be increased or decreased. Temporary workers and the extension of working time are possibilities to increase the production volume. It is conceivable that autonomous teams in a product layout can respond more quickly to the need for additional capacity than functional groups in a process layout (i.e. response flexibility). The need for additional capacity is localized and only one group is involved in the need for more capacity for a particular product family. On the other hand, extension of capacity can be realized more easily in a process layout, because more workers with the same capabilities are eligible to work overtime (i.e. range flexibility). Also, temporary workers can probably best be integrated in a process layout.

The Impact of Layout Type on the Dependability to Deliver on Time

The performance objective "dependability" points to the importance of being dependable with respect to delivery times, the quality of the products, and such. To be dependable, it is important that the manufacturing activities can be buffered from all kinds of disturbances. Machine breakdowns and unexpected absenteeism of workers may complicate the dependability of a manufacturing system. *Interchangeability of machines* (or the ability to subcontract) and the *possibility to replace workers* (or increase the working times of some workers) indicate to what extent a manufacturing system can be reliable in various circumstances. The interchangeability and the possibility to replace workers are probably higher in a process layout because of the clustering of identical capacities. Another aspect of dependability concerns the ability to *control* the flow of products. It is likely that the throughput times in a product layout can be controlled better; the control responsibility can be decentralized to autonomous groups which are able to respond quickly to disturbances.

The Impact of Layout Type on the Quality of Work

Quality of work can be investigated in several ways. A well-known approach concerns the job characteristics model of Hackman and Oldham (1980). This model is used by Huber and Hyer (1985) and Shafer *et* al. (1995) to investigate human issues in cellular manufacturing. The job characteristics model distinguishes five task characteristics that have impact on quality of labor: (i) *skill variety*, (ii) *task identity*, (iii) *task significance*, (iv) *autonomy*, and v) *feedback*. Skill variety refers to the extent to which the work requires a variety of activities involving different skills and talents of the workers. Task identity concerns the extent to which the work enables the worker to complete a whole task from start to finish. Task significance relates to the impact of the work on other people within or outside of the organization. Autonomy indicates to what extent a worker has the freedom to plan, to organize, and to perform the tasks in his/her own way. Finally, feedback refers to the extent to which the worker receives information on the effectiveness of his/her performance.

A product layout likely supports a higher/better skill variety, more autonomy, and a better feedback mechanism: workers can perform a variety of tasks, they are responsible for the internal organization of the group, and they get a quick feedback on their activities. The task identity and task significance in a process layout is probably better: the tasks to be performed are clear for all workers and they will be respected because of their specialization.

DECISION SUPPORT FRAMEWORK

The previous section presented major performance objectives of a manufacturing system and indicated which layout-related aspects play an important role. Table 2 summarizes these aspects. As can be seen, the performance objectives price and speed consist of several elements. These elements together constitute the related performance indicators. We do not distinguish aspects for these two performance objectives.

The ultimate goal is to select the best layout out of a set of alternative layouts. Based upon the scheme of Table 2, the selection problem can be split in three sets of questions:

1. What are the relative scores of the various alternative layouts on the aspects mentioned in Table 2? This question involves a comparison of the alternative layouts with respect to the various aspects. Answering this question requires knowledge of operational issues on the work floor and the ability to assess the impact of an alternative layout on the aspects. The answer to this question determines value $\pi(i,j)$, see Table 3.
2. What is the relative importance of the various aspects for the performance objectives? Table 2 gives an overview of all the aspects. The answer to this question determines value $\pi(j,k)$. The sum of the elements, mentioned in column 3 of Table 2, forms an indication for the performance of respectively the price and speed objective.
3. What is the relative importance of the various performance objectives for the firm? This is basically a strategic question, which has to be answered by the management of the firm. It requires knowledge about customers and competitors. The answer to this question determines value $\pi(k)$.

Table 3. Notation

$\pi(k)$ = relative importance of performance objective k for the firm, $\sum_k \pi(k) = 1$;
$\pi(j,k)$ = relative importance of aspects j on performance objective k, $\sum_j \pi(j,k) = 1$;
$\pi(i,j)$ = relative scores of the layout i on aspect j, $\sum_i \pi(i,j) = 1$;
$R(i)$ = relative performance of alternative *i*, $\sum_i R(i) = 1$.

The answers to these sets of questions enable the calculation of the relative performance of the alternatives:

$$R(i) = \sum_j \sum_k \pi(i,j)\ \pi(j,k)\ \pi(k) \qquad (1)$$

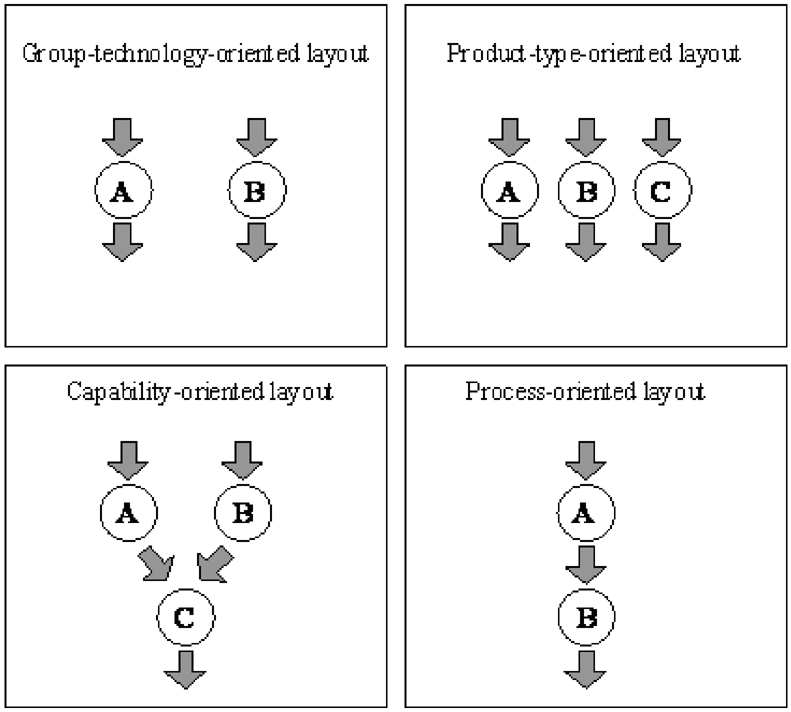
The three sets of questions and the way in which the relative performance of the alternatives are calculated can be seen as an example of using the weighted-score method (see e.g. Slack *et al.* 2001, p. 166). A major issue is the difficulty to "determine" the values of $\pi(k)$, $\pi(j,k)$, $\pi(i,j)$, and $R(i)$. It requires the ability to weight different types of issues.

The three sets of questions and the issue of weighting different types of issues fits in the AHP-approach of Saaty (1980). AHP forces the decision maker(s) to make all assessments explicit. The decomposition of the main problem in several smaller problems also enables an effective participation of employees in the decision problem, using their specific expertise and responsibilities. In the next section, we will illustrate the use of the AHP-approach for layout selection on hand of a practical instance.

CASE STUDY

The case study presented here concerns the sheet metal processing department of the firm Holec Algemene Toelevering B.V., a supplier of parts, tools, and services for the electro-mechanical industry. Before the layout study started, the sheet metal processing department consisted of four autonomous manufacturing cells with some exchangeability between the cells: (i) an automated flexible system for sheet metal working (<3 mm), (ii) numerical sheet metal working (>3mm), (iii) sheet metal construction processing (>5 mm), and (iv) conventional sheet metal processing. Basic processes to be performed in the cells are sawing, punching, cutting, tapping, squaring, welding, and bench work. The firm started to produce in manufacturing cells in 1987. This has led to significant improvements in manufacturing throughput time and efficiency. In the course of years, however, there were several reasons to move back to a more functional layout, such as the complexity and productivity of new equipment, the possibility of workers to operate more than one machine simultaneously, and the increased variety of part types. Other parts of the manufacturing facility of the firm were already transformed to a more process-oriented layout (see Molleman *et al.* 2002). A layout study at the sheet metal processing department started in 2000. Four alternative layouts were generated on the basis of a production flow analysis (Burbidge 1991, Slomp 1998): (i) a group-technology-oriented alternative, (ii) a product-type-oriented alternative, (iii) a capability-oriented alternative, and (iv) a process-oriented alternative. These alternatives are schematically depicted in Figure 3.

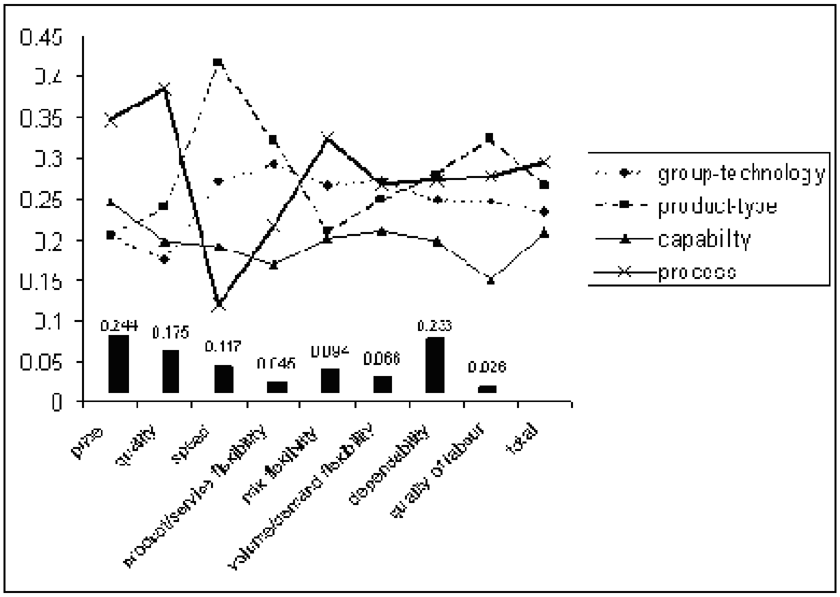
Figure 3. Layout type alternatives for the sheet metal processing



In the group-technology-oriented alternative, the department is divided in two relatively autonomous cells with minimal intercell movements. Some part types can be produced in both groups, which simplifies the balancing of the workload. The product-type-oriented alternative consists of three manufacturing cells, each responsible for a particular type of product. Two cells are responsible for the production of repetitive part types, while one cell is mainly focused on the production of quick orders. An important advantage of this layout is that the production of repetitive part types is not disturbed by quick orders. This may simplify the production control. On the other hand, the cell that is responsible for the quick orders may face undesirable fluctuations of demand. The capability-oriented alternative consists of three manufacturing cells. Basic viewpoint of this alternative is that all assembly work (welding and bench work) needs to be performed in one manufacturing cell (C). The other cells (A and B) are autonomous cells, which have their own product-oriented capabilities. Therefore, intercell movements are minimal. The process-oriented alternative consists of two manufacturing cells. Sawing, punching and cutting is performed in cell A, while tapping, squaring, welding, and bench work is done in cell B. Each cell consists of small groups of identical machines.

The four alternatives are compared by means of the decision framework of Table 2 and by using the AHP methodology. We used the software package Expert Choice. Figure 4 presents the results of the comparisons of the four alternatives. The bars in Figure 4 indicate the relative importance of the various performance objectives. The four lines in the figure show the relative scores of each alternative on the performance objectives. The position of the alternatives at “total” shows the final judgment of the alternatives. As can be seen, the process-oriented layout is preferred because of its positive effect on price, quality,

Figure 4. Scores of the four alternatives



mix flexibility and, to less extent, dependability. Especially price and quality are important reasons to select the process-oriented layout.

It is interesting to see the almost equal end scores of the product-type-oriented and the process-oriented layout, despite their completely different orientation. The product-type-oriented layout performs well with respect to the performance objective speed. In the assessment of the speed factors of Table 2, the managers of the firm assumed that the production control in the process-oriented layout is more complex and will perform worse than in the product-type-oriented layout. At that moment, the firm did not have a good registration system (bar-coding system) on the work floor that is connected with the production control system. A better shop floor control system, which was under study at the firm at the moment of deciding for a new manufacturing layout, would likely improve the score of the process-oriented layout on the performance objective speed. This kind of sensitivity analysis is also useful for the assessment of the scores on the performance objectives quality and price. In this particular case, the impact of shorter control loops on the quality of the products is assessed as being minimal. This assessment has a negative impact on the final score of the product-type-oriented layout. The software package Expert Choice supports sensitivity analysis. It appears that if short control loops do have a major impact on the quality of the products, the product-type-oriented layout performs better than the process-oriented alternative.

Based upon the results of the analysis, the firm changed the layout of the sheet-metal processing department into a process-oriented layout, see also Molleman *et al.* (2002). The systematic approach of the selection problem is seen at the firm as a major help to canalize the discussions about the required layout of manufacturing departments.

CONCLUSION AND REFLECTIONS

This chapter presents a systematic approach for the selection of a manufacturing layout type. The approach includes the use of the AHP-methodology. An important element of the approach is the construction of a decision hierarchy and the pairwise comparisons of decision elements. In this section we will first reflect on the AHP methodology and next we will draw conclusions on the use of AHP for the layout selection type problem.

As in all Multi-Criteria-Decision-Methods, AHP is sensitive for issues such as the specification of the selection problem, its decomposition, and the scales used for the pairwise comparisons (see Pöyhönen *et al.* 1997). The quality of the outcome of an AHP analysis is largely determined by the quality of the problem specification. For instance, adding aspects or regrouping decision elements may lead to different outcomes. A particular problem concerns the issue of "rank reversal". This means that the priority of alternatives may change if alternatives are removed from and/or other alternatives are added to the selection problem (see e.g. Belton and Gear 1983). The problem of rank reversal plays a role if almost identical alternatives are taken into consideration. Finally, the number of pairwise comparisons may be problematic and may lead to unreliable results. Employees who have to make the pairwise comparisons may get tired and lose the required concentration. Another issue, which has to be taken into consideration when applying the AHP approach, is the translation of verbal or graphical assessments in numerical figures. Pöyhönen *et al.* (1997) show that it is not advisable to mix different types of assessment within the levels of the AHP hierarchy.

This chapter has presented a decision support framework based on the AHP approach for the selection of a manufacturing layout. The value of the framework is illustrated by means of a case application. Important advantages of using the AHP approach are (1) the ability to decompose the complex decision problem in smaller problems, (2) the possibility of an efficient and effective employee participation, and (3) the detailed assessment of the selected layout alternative, which helps to define further improvement actions. These advantages of using the AHP approach are also illustrated by Abdi and Labid (2003).

Interesting point in the case study, as presented in this chapter, is that opposite alternatives do have the best scores. This illustrates the group technology debate, as it takes place in practice. Both alternatives appear to be acceptable and have their pros and cons. The proposed approach has the advantage that it gives insight in whether the two alternatives do have similar scores. A debate about the differences on the scores of the various aspects will help the decision process and the acceptability of the final decision.

The systematic approach as presented in this chapter is developed around 1999 (see also Slomp *et al.* 1999a, b) and is applied in several practical situations, mostly master projects of Industrial Engineering students. The evaluation criteria (see Table 2) and the use of AHP has proven to be a robust framework for the selection of a layout in many situations.

REFERENCES

Abdi, M. R., & Labib, A. W. (2003). A design strategy for reconfigurable manufacturing systems (RMSs) using analytical hierarchical process (AHP): A case study. *International Journal of Production Research*, *41*(10), 2273–2299. doi:10.1080/0020754031000077266

Agarwal, A., & Sarkis, J. (1998). A review and analysis of comparative performance studies on functional and cellular manufacturing layouts. *Computers & Industrial Engineering*, *34*(1), 77–89. doi:10.1016/S0360-8352(97)00152-6

Balakrishnan, J., & Cheng, C. H. (1998). The dynamic plant layout problem—A state-of-the-art review. *Omega*, *26*(4), 507–521. doi:10.1016/S0305-0483(97)00078-9

Belton, V., & Gear, T. (1983). On a short-coming of Saaty's method of analytic hierarchies. *Omega*, *11*(3), 228–230. doi:10.1016/0305-0483(83)90047-6

Benjaafar, S., Heragu, S. S., & Irani, S. A. (2002). Next generation factory layouts: Research challenges and recent progress. *Interfaces*, *32*(6), 58–77. doi:10.1287/inte.32.6.58.6473

Benjaafar, S., & Sheikhzadeh, M. (2000). Design of flexible layouts. *IIE Transactions*, *32*, 309–322. doi:10.1080/07408170008963909

Burbidge, J. L. (1992). Change to group technology: process organization is obsolete. *International Journal of Production Research*, *30*(5), 1209–1219. doi:10.1080/00207549208942951

Burbidge, J. L., Partridge, J. T., & Aitchison, K. (1991). Planning group technology for Davy Morris using production flow analysis. *Production Planning and Control*, *2*(1), 59–72. doi:10.1080/09537289108919331

Cambron, K. E., & Evans, G. W. (1991). Layout design using the analytic hierarchy process. *Computers & Industrial Engineering*, *20*(2), 211–229. doi:10.1016/0360-8352(91)90026-3

Ertay, T., Ruan, D., & Tuzkaya, U. R. (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing system. *Information Sciences*, *176*, 237–262. doi:10.1016/j.ins.2004.12.001

Farrington, P. A., & Nazemetz, J. W. (1998). Evaluation of the performance domain of cellular and functional layouts. *Computers & Industrial Engineering*, *34*(1), 91–101. doi:10.1016/S0360-8352(97)00153-8

Flynn, B. B., & Jacobs, F. R. (1986). A simulation comparison of group technology with traditional job shop manufacturing. *International Journal of Production Research*, *24*(5), 1171–1192. doi:10.1080/00207548608919795

Francis, R. L., McGinis, L. F., & White, J. A. (1992). *Facility layout and location—An analytical approach*. Englewood Cliffs, NJ: Prentice Hall.

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.

Huber, V. L., & Hyer, N. L. (1985). The human factor in cellular manufacturing. *Journal of Operations Management*, *5*(2), 213–228. doi:10.1016/0272-6963(85)90008-7

Irani, S. A., Cavalier, T. M., & Cohen, P. H. (1993). Virtual manufacturing cells: Exploiting layout design and intercell flows for the machine sharing problem. *International Journal of Production Research*, *31*(4), 791–810. doi:10.1080/00207549308956757

Irani, S. A., & Huang, H. (2000). Custom design of facility layouts for multi-product facilities using layout modules. *IEEE Transactions on Robotics and Automation*, *16*, 259–267. doi:10.1109/70.850644

Johnson, D. J., & Wemmerlöv, U. (1996). On the relative performance of functional and cellular layouts – An analysis of the model-based comparative studies literature. *Production and Operations Management*, *5*(4), 309–334. doi:10.1111/j.1937-5956.1996.tb00403.x

Labib, A. W., Williams, G. B., & O'Connor, R. F. (1998). An intelligent maintenance model (system): An application of the analytic hierarchy process and a fuzzy logic rule-based controller. *Operational Research Society*, *49*, 745–757.

Meller, R. D., & Gau, Y. K. (1996). The facility layout problem: Recent and emerging trends, and perspectives. *Journal of Manufacturing Systems*, *15*, 351–366. doi:10.1016/0278-6125(96)84198-7

Meyers, F. E., & Stephens, M. P. (2004). *Manufacturing facilities design and material handling* (3rd ed.). Prentice Hall.

Molleman, E., Slomp, J., & Rolefes, S. (2002). The evolution of a cellular manufacturing system - A longitudinal case study. *International Journal of Production Economics*, *75*(3), 305–322. doi:10.1016/S0925-5273(01)00139-6

Montreuil, B., LeFrancois, P., Marcotte, S., & Venkatadri, U. (1993). *Holographic layout of manufacturing systems operating in chaotic environments.* Technical Report 93-53, Document de Recherche GRGL, Faculte des Sciences de l'Administration, Universite Laval, Quebec.

Montreuil, B., Venkatadri, U., & Rardin, R. L. (1999). Fractal layout organization for job shop environments. *International Journal of Production Research*, *37*(3), 501–521. doi:10.1080/002075499191643

Muther, R. (1973). *Systematic layout planning* (2nd ed.). Boston, MA: Cahners Books.

Partove, F. Y., & Burton, J. (1992). An analytic hierarchy approach to facility layout. *Computers & Industrial Engineering*, *22*(4), 447–457. doi:10.1016/0360-8352(92)90020-K

Pöyhönen, M., Hämäläinen, R. P., & Salo, A. A. (1997). An experiment on the numerical modeling of verbal ratio statements. *Journal of Multi-Criteria Decision Analysis*, *6*, 1–10. doi:10.1002/(SICI)1099-1360(199701)6:1<1::AID-MCDA111>3.0.CO;2-W

Raman, D., Nagalingam, S. V., & Lin, G. C. I. (2009). Towards measuring the effectiveness of a facilities layout. *Robotics and Computer-integrated Manufacturing*, *25*, 191–203. doi:10.1016/j.rcim.2007.06.003

Rosenblatt, M. J. (1986). The dynamics of plant layout. *Management Science*, *32*(1), 76–86. doi:10.1287/mnsc.32.1.76

Saaty, T. L. (1980). *The analytic hierarchy process*. New York, NY: McGraw-Hill.

Shafer, S. M., & Charnes, J. M. (1993). Cellular versus functional layouts under a variety of shop operating conditions. *Decision Sciences*, *24*(3), 665–681. doi:10.1111/j.1540-5915.1993.tb01297.x

Shafer, S. M., & Meredith, J. R. (1993). An empirically-based simulation study of functional versus cellular layouts with operations overlapping. *International Journal of Operations & Production Management*, *13*(2), 47–62. doi:10.1108/01443579310025303

Shafer, S. M., Tepper, B. J., Meredith, J. R., & Marsh, R. (1995). Comparing the effects of cellular and functional manufacturing on employees' perceptions and attitudes. *Journal of Operations Management*, *12*, 63–74. doi:10.1016/0272-6963(94)00011-3

Shambu, G., Suresh, N. C., & Pegels, C. C. (1996). Performance evaluation of cellular manufacturing systems: A taxonomy and review of research. *International Journal of Operations & Production Management*, *16*(8), 81–103. doi:10.1108/01443579610125796

Slack, N. (1987). The flexibility of manufacturing systems. *International Journal of Operations & Production Management*, *7*(4), 35–45. doi:10.1108/eb054798

Slack, N., Chambers, S., & Johnston, R. (2001). *Operations management.* Harlow, MA: Financial Times/Prentice Hall.

Slomp, J. (1998). Design of manufacturing cells: PFA applications in Dutch industry. In Suresh, N. C., & Kay, J. M. (Eds.), *Group technology and cellular manufacturing: A state-of-the-art synthesis of research and practice* (pp. 153–168). Boston, MA: Kluwer Academic. doi:10.1007/978-1-4615-5467-7_9

Slomp, J., Molleman, E., & Gaalman, G. J. C. (1993). Production and operations management aspects of cellular manufacturing – a survey of users. In I. A. Pappas & I. P. Tatsiopoulos (Ed.), *Advances in production management systems* (pp. 553-560). North-Holland: Elsevier Science Publishers B.V.

Slomp, J., Molleman, E., & Suresh, N. C. (1995). Empirical investigation of the performance of cellular manufacturing, *Proceedings Management and New Production Systems, EUROMA conference*, (pp. 432-441).

Slomp, J., Vrolijk, H., & Bokhorst, J. (1999). [a]). Beoordeling van alternatieve productiestructuren (deel 1). [in Dutch]. *MB-dossier*, *65*(6), 6–11.

Slomp, J., Vrolijk, H., & Bokhorst, J. (1999). [b]). AHP als methode voor de keuze van een geschikte productiestructuur - Gestructureerde besluitvorming in de keuze tussen alternatieve structuren (deel 2). [in Dutch]. *MB-dossier*, *65*(7/8), 6–12.

Suresh, N. C. (1991). Partitioning work centers for group technology: insights from an analytical model. *Decision Sciences*, *22*(4), 772–791. doi:10.1111/j.1540-5915.1991.tb00364.x

Suresh, N. C., & Meredith, J. R. (1994). Coping with the loss of pooling synergy in cellular manufacturing systems. *Management Science*, *40*(4), 466–483. doi:10.1287/mnsc.40.4.466

Upton, D. M. (1994). The management of manufacturing flexibility. *California Management Review*, *36*(2), 72–89.

Venkatadri, U., Rardin, R. L., & Montreuil, B. (1997). A design methodology for fractal layout organization. *IIE Transactions*, *29*, 911–924. doi:10.1080/07408179708966411

Wemmerlöv, U., & Hyer, N. L. (1989). Cellular manufacturing in the U.S. industry: A survey of users. *International Journal of Production Research*, *27*(9), 1511–1530. doi:10.1080/00207548908942637

Wemmerlöv, U., & Johnson, D. J. (1997). Cellular manufacturing at 46 user plants: Implementation experiences and performance improvements. *International Journal of Production Research*, *35*(1), 29–49. doi:10.1080/002075497195966

Yang, T., & Hung, C.-C. (2007). Multiple-attribute decision making methods for plant layout design problem. *Robotics and Computer-integrated Manufacturing*, *23*, 126–137. doi:10.1016/j.rcim.2005.12.002

Yang, T., & Kuo, C. (1997). A hierarchical AHP/DEA methodology for the facilities layout design problem. *European Journal of Operational Research, 147*(2003), 128–136

Zahedi, F. (1986). The analytic hierarchy process – A survey of the method and its applications. *Interfaces*, *16*, 96–108. doi:10.1287/inte.16.4.96

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 21-36, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 23

Petri Net Model Based Design and Control of Robotic Manufacturing Cells

Gen'ichi Yasuda
Nagasaki Institute of Applied Science, Japan

ABSTRACT

The methods of modeling and control of discrete event robotic manufacturing cells using Petri nets are considered, and a methodology of decomposition and coordination is presented for hierarchical and distributed control. Based on task specification, a conceptual Petri net model is transformed into the detailed Petri net model, and then decomposed into constituent local Petri net based controller tasks. The local controllers are coordinated by the coordinator through communication between the coordinator and the controllers. Simulation and implementation of the control system for a robotic workcell are described. By the proposed method, modeling, simulation, and control of large and complex manufacturing systems can be performed consistently using Petri nets.

INTRODUCTION

Manufacturing systems, where the materials which are handled are mainly composed of discrete entities, for example parts that are machined and/or assembled, are called discrete manufacturing systems. Due to its complexity, manufacturing system control is commonly decomposed into a hierarchy of abstraction levels: planning, scheduling, coordination and local control. Each level operates on a certain time horizon. The planning level determines at which time each product will be introduced in the manufacturing system. The scheduling level produces a sequence of times for the execution of each operation on each machine or a total ordering of all the operations. The coordination level updates the state representation of the manufacturing system in real-time, supervises it and makes real-time decisions. The local control level implements the real-time

DOI: 10.4018/978-1-4666-1945-6.ch023

control of machines and devices etc., interacting directly with the sensors and actuators. All the emergency procedures are implemented at this level, so real-time constraints may be very hard. At each level, any modeling has to be based on the concepts of discrete events and states, where an event corresponds to a state change (Martinez, 1986), (Silva, 1990).

A flexible manufacturing system is formed of a set of flexible machines, an automatic transport system, and a sophisticated decision making system to decide at each instant what has to be done and on which machine. A manufacturing cell is an elementary manufacturing system consisting of some flexible machines (machine tools, assembly devices, or any complex devices dedicated to complex manufacturing operations), some local storage facilities for tools and parts and some handling devices such as robots in order to transfer parts and tools. Elementary manufacturing cells are called workstations. At the local control level of manufacturing cells many different kinds of machines can be controlled, and specific languages for different application domains are provided; for example, block diagrams for continuous process control and special purpose languages for CNC or robot programming. For common sequential control, special purpose real-time computers named Programmable Logic Controllers (PLCs) are used. PLCs are replacements for relays, but they incorporate many additional and complex functions, such as supervisory and alarm functions and start-up and shut-down operations, approaching the functionalities of general purpose process computers. The most frequent programming languages are based on ladder or logic diagrams and boolean algebra. However, when the local control is of greater complexity, the above kinds of languages may not be well adapted. The development of industrial techniques makes a sequential control system for manufacturing cells more large and complicated one, in which some subsystems operate concurrently and cooperatively. Conventional representation methods based on flowcharts, time diagrams, state machine diagrams, etc. cannot be used for such systems.

To realize control systems for flexible manufacturing cells, it is necessary to provide effective tools for describing process specifications and developing control algorithms in a clear and consistent manner. In the area of real-time control of discrete event manufacturing cells the main problems that the system designer has to deal with are concurrency, synchronization, and resource sharing problems. For this class of problems, Petri nets have intrinsic favorable qualities and it is very easy to model sequences, choices between alternatives, rendezvous and concurrent activities by means of Petri nets (Reisig, 1985). When using Petri nets, events are associated with transitions. Activities are associated to the firing of transitions and to the markings of places which represent the states of the system. The network model can describe the execution order of sequential and parallel tasks directly without ambiguity (Murata, et al. 1986), (Crockett, et al. 1987). Moreover, the formalism allowing a validation of the main properties of the Petri net control structure (liveness, boundedness, etc.) guarantees that the control system will not fall immediately in a deadlocked situation. In the field of flexible manufacturing cells, the last aspect is essential because the sequences of control are complex and change very often. Furthermore, a real-time implementation of the Petri net specification by software called a token player can avoid implementation errors, because the specification is directly executed by the token player and the implementation of these control sequences preserves the properties of the model (Bruno, 1986). In this approach, the Petri net model is stored in a database and the token player updates the state of the database according to the operation rules of the model. For control purposes, this solution is very well suited to the need of flexibility, because, when the control sequences change, only the database needs to be changed(Silva, et al. 1982), (Valette, et al. 1983).

In addition to its graphic representation differentiating events and states, Petri nets allows the modeling of true parallelism and the possibility of progressive modeling by using stepwise refinements or modular composition. Libraries of well-tested subnets allow components reusability leading to significant reductions in the modeling effort. The possibility of progressive modeling is absolutely necessary for flexible manufacturing cells because they are usually large and complex systems. The refinement mechanism allows the building of hierarchically structured net models.

Some techniques derived from Petri nets have been successfully introduced as an effective tool for describing control specifications and realizing the control in a uniform manner. However, in the field of flexible manufacturing cells, the network model becomes complicated and it lacks the readability and comprehensibility. Therefore, the flexibility and expandability are not satisfactory in order to deal with the specification change of the control system. Despite the advantages offered by Petri nets, the synthesis, correction, updating, etc. of the system model and programming of the controllers are not simple tasks (Desrochers, et al. 1995), (Lee, et al. 2006).

In this chapter, a Petri net based specification and real-time control method for large and complex manufacturing cells is presented. Based on the hierarchical and distributed structure of the manufacturing cell, the specification procedure is a top-down approach from the conceptual level to the detailed level such that the macro representation of the system is broken down to generate the detailed Petri nets at the local machine control level. Then the Petri nets are decomposed and assigned to the machine controllers to perform distributed control using Petri net based multitask processing. An algorithm is proposed for coordination of machine controllers. By the proposed method, modeling, simulation and control of large and complex manufacturing cells can be performed consistently using Petri nets.

MODELING OF MANUFACTURING CELLS USING MODIFIED PETRI NETS

A manufacturing process is characterized by the flow of workpieces or parts, which pass in ordered form through subsystems and receive appropriate operations. Each subsystem executes manufacturing operations, that is, physical transformations such as machining, assembling, or transfer operations such as loading and unloading.

From the viewpoint of discrete event process control, an overall manufacturing process can be decomposed into a set of distinct activities (or events) and conditions mutually interrelated in a complex form. An activity is a single operation of a manufacturing process executed by a subsystem. A condition is a state in the process such as machine operation mode.

Considering the nature of discrete event manufacturing cells which are characterized by the occurrence of events and changing conditions, the condition-event net based specification method has been investigated. The Petri net is one of the effective means to represent condition-event systems. The specification method is a graphical model used as a tool to identify types of activities, conditions, and their mutual interrelation. It describes explicitly the concept of the manufacturing process to be carried out in the discrete event manufacturing cells.

Considering not only the modeling of the systems but also the actual well-designed control, the guarantee of safeness and the capability to represent input and output signals from and to the machines are required. In the condition-event systems, deadlock occurs when the system enters into a state that is not possible for any event to occur. Further bumping occurs when, despite the holding of a condition, the preceding event occurs. This can result in the multiple holding of that condition. Therefore, the basic Petri net which is called Place/Transition-net should be modified and extended in order to represent the activity contents and control strategies for the manufacturing system control in detail.

The extended Petri net consists of the following six elements: (1) Place, (2) Transition, (3) Directed arc, (4) Token, (5) Gate arc, (6) Output signal arc (Hasegawa, et al. 1984). A place represents a condition of a system element or action. A transition represents an event of the system. A directed arc connects from a place to a transition or from a transition to a place, and its direction shows the input and output relation between them. Places and transitions are alternately connected using directed arcs. The number of directed arcs connected with places or transitions is not restricted. A token is placed in a place to indicate that the condition corresponding to the place is holding.

A gate arc connects a transition with a signal source, and depending on the signal, it either permits or inhibits the occurrence of the event which corresponds to the connected transition. Gate arcs are classified as permissive or inhibitive, and internal or external. An output signal arc sends the signal from a place to an external machine. Thus a transition is enabled if and only if it satisfies all the following conditions:

1. It does not have any output place filled with a token.
2. It does not have any empty input place.
3. It does not have any internal permissive arc signaling 0.
4. It does not have any internal inhibitive arc signaling 1.

Figure 1 shows the place and gate variables for transition firing test.

Formally, the enabling condition and the external gate condition of a transition j are described using the logical place and gate variables as follows:

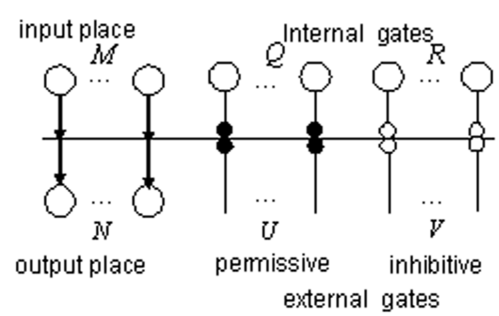
$$t_j(k) = \bigcap_{m=1}^{M} p_{j,m}^{I}(k) \wedge \bigcap_{n=1}^{N} \overline{p_{j,n}^{O}(k)} \wedge \bigcap_{q=1}^{Q} g_{j,q}^{IP}(k) \wedge \bigcap_{r=1}^{R} \overline{g_{j,r}^{II}(k)} \quad (1)$$

$$g_j^{E}(k) = \bigcap_{u=1}^{U} g_{j,u}^{EP}(k) \wedge \bigcap_{v=1}^{V} \overline{g_{j,v}^{EI}(k)} \quad (2)$$

where,

- M: set of input places of transition j
- $p_{j,m}^{I}(k)$: state of input place m of transition j at time sequence k
- N: set of output places of transition j
- $p_{j,n}^{O}(k)$: state of output place n of transition j at time sequence k
- Q: set of internal permissive gate signals of transition j
- $g_{j,q}^{IP}(k)$: internal permissive gate signal variable q of transition j at time sequence k
- R: set of internal inhibitive gate signals of transition j
- $g_{j,r}^{II}(k)$: internal inhibitive gate signal variable r of transition j at time sequence k
- U: set of external permissive gate signals of transition j
- $g_{j,u}^{EP}(k)$: external permissive gate signal variable u of transition j at time sequence k

Figure 1. Place and gate variables for transition firing test



- V: set of external inhibitive gate signals of transition j
- $g_{j,v}^{EI}(k)$: external inhibitive gate signal variable v of transition j at time sequence k

The state (marking) change, that is, the addition or removal of a token of an input or output place, is described as follows:

$$p_{j,m}^{I}(k+1) = p_{j,m}^{I}(k) \wedge \overline{(t_j(k) \wedge g_j^E(k))} \qquad (3)$$

$$p_{j,n}^{O}(k+1) = p_{j,n}^{O}(k) \vee (t_j(k) \wedge g_j^E(k)) \qquad (4)$$

An enabled transition may fire when it does not have any external permissive arc signaling 0 nor any external inhibitive arc signaling 1. The firing of a transition removes a token from each input place and put a token in each output place connected to it. The assignment of tokens into the places of a Petri net is called marking and it represents the system state. In any initial marking, there must not exist more than one token in a place. According to these rules, the number of tokens in a place never exceeds one, thus the Petri net is essentially a safe graph; the system is free from the bumping phenomenon.

For the actual control, the operations of each machine are broken down into a series of unit motions, which is represented by mutual connection between places and transitions. A place means a concrete unit motion of a machine. From these places, output signal arcs are connected to the machines, and external gate arcs from the machines are connected to the transitions of the Petri net when needed, for example, to synchronize and coordinate operations. When a token enters a place that represents a subtask, the machine defined by a machine code is informed to execute a specified subtask with positional data and control parameters; all the code and data are defined as the place parameters.

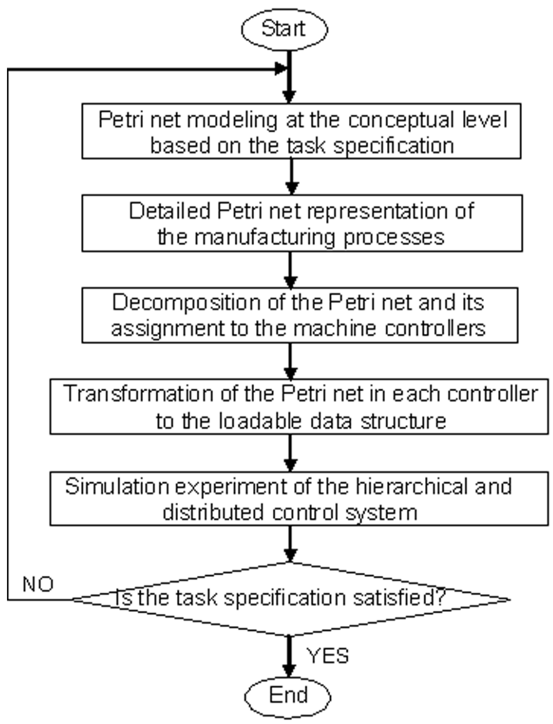
If a place has two or more input or output transitions, these transitions may be in conflict for firing. When two or more transitions are enabled only one transition should fire using gate arcs or some arbitration rule. The Petri net described in detail by such a procedure mentioned above can be used as a program for the system control, while features of discrete event manufacturing cells such as ordering, parallelism, asynchronism, concurrency and conflict can be concretely described through the extended Petri net.

The extended Petri net is a tool for the study of condition-event systems and used to model condition-event systems through its graphical representation. Analysis of the net reveals important information about the structure and the dynamic behavior of the modeled condition-event system. This information can then be used to evaluate the modeled condition-event system and suggest improvements or changes.

DESIGN OF HIERARCHICAL AND DISTRIBUTED CONTROL SYSTEM

A specification procedure for discrete event manufacturing cells based on Petri nets is as follows. First, the conceptual level activities of the discrete event manufacturing cells are defined through a Petri net model considering the task specification corresponding to the manufacturing process. A conceptual Petri net model describes the aggregate manufacturing process. At the level, each subtask composing the task specification is represented as a place of the Petri net, where the activity of each equipment is also represented as a place. Then, the detailed Petri nets describing the activities are deduced based on activity specification and required control strategies. The macro representation of the manufacturing process is effectively used for achieving a top-down interpretation down to the concrete lower level activities using Petri nets. Based on the hierarchical approach, the Petri net is translated into the detailed Petri net by

Figure 2. Flow chart of Petri net based implementation of hierarchical and distributed control system



stepwise refinements from the highest conceptual level to the lowest machine control level. This procedure is repeated up to an appropriate level corresponding to the control level of the equipment responsible for the activity execution. At each step of detailed specification, places of the Petri net are substituted by a subnet in a manner which maintains the structural properties (Miyagi, 1988). The overall procedure of the Petri net based implementation of hierarchical and distributed control for robotic manufacturing cells is summarized as shown in Figure 2.

It is natural to implement a hierarchical and distributed control system, where one controller is allocated to each control layer or block. For the robotic manufacturing cells composed of robots, machine tools, and conveyors, an example structure of hierarchical and distributed control is composed of one station controller and three machine controllers as shown in Figure 3, although each robot may be controlled by one robot controller. The detailed Petri net is decomposed into subnets, which are assigned to machine controllers.

In the decomposition procedure, a transition may be divided and distributed into different machine controllers as shown in Figure 4. The machine controllers should be coordinated so that these transitions fire simultaneously, that is, the aggregate behavior of decomposed subnets should be the same as that of the original Petri net. Decomposed transitions are called global transitions, and other transitions are called local transitions.

By the Petri net model, the state of the discrete event system is represented as the marking of tokens, and firing of any transition brings about change to the next state. So the firing condition and state (marking) change before decomposition should be the same as those after decomposition. If transition j is divided into s transitions $j1$, $j2$,,,

Figure 3. Example structure of distributed control system

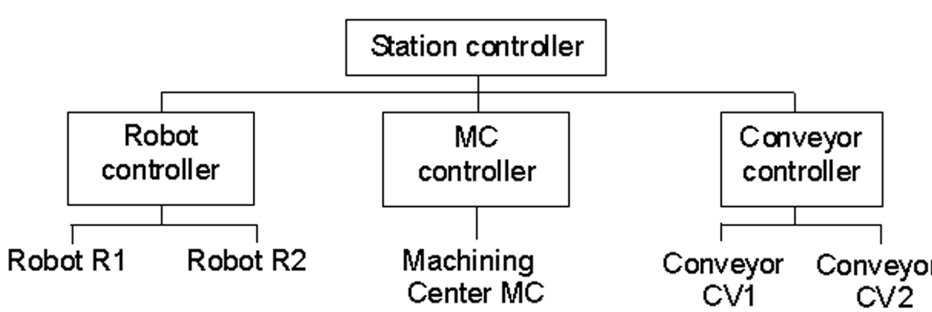
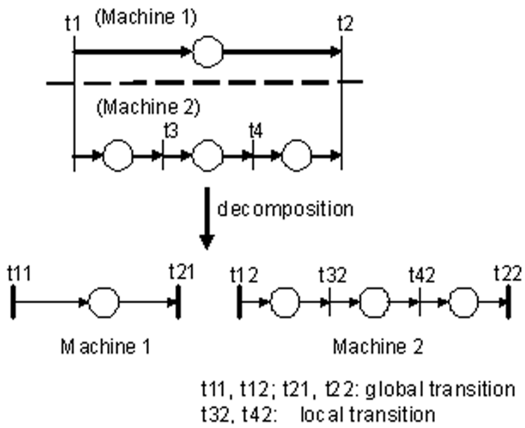


Figure 4. Decomposition of transitions



js, as shown in Figure 5, the firability condition of the transition after decomposition is described as follows:

$$t_{jsub}(k) = \bigcap_{m=1}^{Msub} p^{I}_{jsub,m}(k) \wedge \bigcap_{n=1}^{Nsub} \overline{p^{O}_{jsub,n}(k)} \wedge \bigcap_{q=1}^{Qsub} g^{IP}_{jsub,q}(k) \wedge \bigcap_{r=1}^{Rsub} \overline{g^{II}_{jsub,r}(k)} \tag{5}$$

$$g^{E}_{jsub}(k) = \bigcap_{u=1}^{Usub} g^{EP}_{jsub,u}(k) \wedge \bigcap_{v=1}^{Vsub} \overline{g^{EI}_{jsub,v}(k)} \tag{6}$$

From Equation 1 and Equation 5,

$$t_{j}(k) = \bigcap_{sub=1}^{S} t_{jsub}\ (k) \tag{7}$$

From Equation 2 and Equation 6,

$$g^{E}_{j}(k) = \bigcap_{sub=1}^{S} g^{E}_{jsub}(k) \tag{8}$$

where,

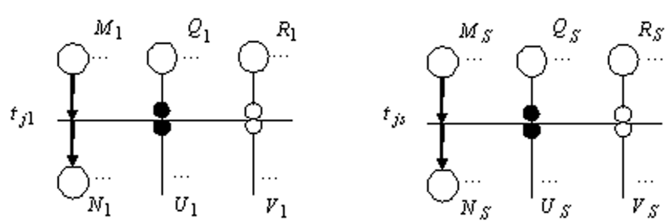
- *S*: total number of subnets
- *Msub*: set of input places of transition *jsub* of subnet *sub*
- $p^{I}_{jsub,m}(k)$: state of input place *m* of transition *jsub* of subnet *sub* at time sequence *k*
- *Nsub*: set of output places of transition *jsub* of subnet *sub*
- $p^{O}_{jsub,n}(k)$: state of output place *n* of transition *jsub* of subnet *sub* at time sequence *k*
- *Qsub*: set of internal permissive gate signals of transition *jsub* of subnet *sub*
- *Rsub*: set of internal inhibitive gate signals of transition *jsub* of subnet *sub*
- *Usub*: set of external permissive gate signals of transition *jsub* of subnet *sub*
- *Vsub*: set of external permissive gate signals of transition *jsub* of subnet *sub*

The addition or removal of a token of a place connected to a decomposed transition is described as follows:

$$p^{I}_{jsub,m}(k+1) = p^{I}_{jsub,m}(k) \wedge \overline{(t_{j}(k) \wedge g^{E}_{j}(k))} \tag{9}$$

$$p^{O}_{jsub,n}(k+1) = p^{O}_{jsub,n}(k) \vee (t_{j}(k) \wedge g^{E}_{j}(k)) \tag{10}$$

Figure 5. Place and gate variables after decomposition of transition

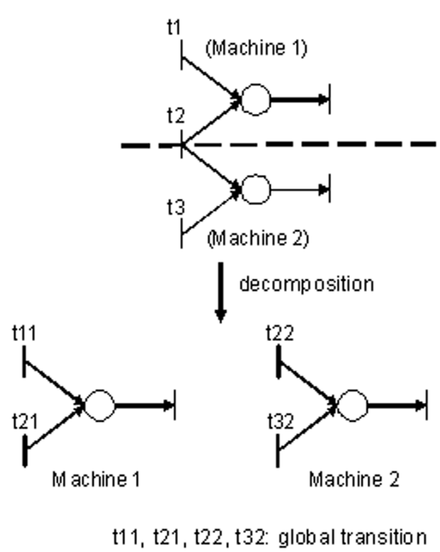


Consequently it is proven that the firability condition of the original transition is equal to AND operation of firability conditions of decomposed transitions. If and only if all of the decomposed transitions are firable, then the global transitions are firable. To utilize the above results, the coordinator program has been introduced to coordinate the decomposed subnets so that the aggregate behavior of decomposed subnets is the same as that of the original Petri net.

There may exist a place which has several input transitions and/or several output transitions. This place is called a conflict place. The transitions connected to a conflict place are in conflict when some of them are firable at the same time. In this case, only one of them can be fired and the others become disabled. The selection of firing transition is done arbitrarily using an arbiter program.

In case that a transition in conflict with other transitions is decomposed as shown in Figure 6, these transitions should be coordinated by the system controller. If arbitrations of the transitions are performed independently in separate subnets, the results may be inconsistent with the original rule of arbitration. Therefore the transitions should be arbitrated together as a group. On the other hand, arbitration of local transitions in conflict is performed by local machine controllers.

Figure 6. Decomposition of transition in conflict

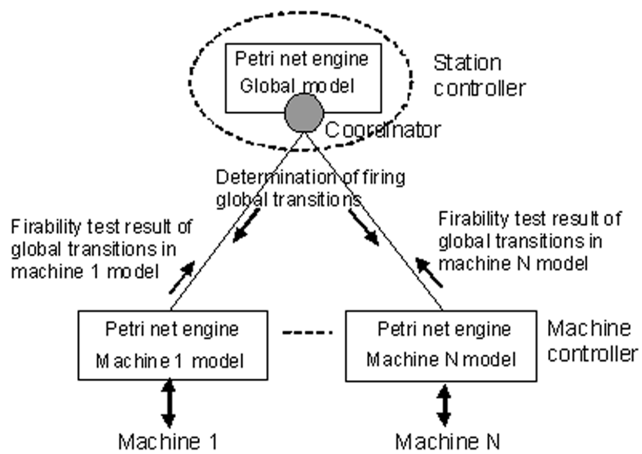


The hierarchical and distributed control system composed of one station controller and several machine controllers has been implemented. The conceptual Petri net model is allocated to the Petri net based controller for management of the overall system. The detailed Petri net models are allocated to the Petri net based controllers in the machine controllers. Each machine controller directly monitors and controls the sensors and actuators of its machine.

The control of the overall system is achieved by coordinating these Petri net based controllers. Figure 7 shows the Petri net based control structure with the coordinator. System coordination is performed through communication between the coordinator in the station controller and the Petri net based controllers in the machine controllers as the following steps.

1. When each machine controller receives the start signal from the coordinator, it tests the firability of all transitions in its own Petri net, and sends the information on the global transitions and the end signal to the coordinator.
2. The coordinator tests the firability of the global transitions, arbitrates conflicts among global transitions, and sends the names of firing global transitions and the end signal to the machine controllers.
3. Each machine controller arbitrates conflicts among local transitions using the information from the coordinator, generates a new marking, and sends the end signal to the coordinator.
4. When the coordinator receives the end signal from all the machine controllers, it sends the output command to the machine controllers.
5. Each machine controller outputs the control signals to its actuators simultaneously.

Figure 7. Petri net based control structure with coordinator

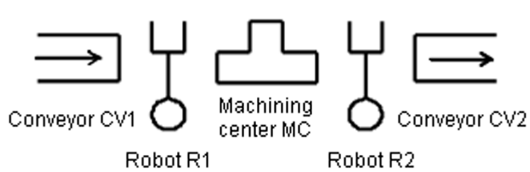


REAL-TIME CONTROL OF A MANUFACTURING CELL

The example manufacturing system has two robots, one machining center, and two conveyors, where one is for carrying in and the other is for carrying out, as shown in Figure 8. The main execution of the system is indicated as the following task specification:

1. A workpiece is carried in by the conveyor CV1.
2. The robot R1 loads the workpiece to the machining center MC.
3. The machining center MC processes the workpiece.
4. The robot R2 unloads the workpiece from the machining center and places it on the conveyor CV2.

Figure 8. Example of robotic manufacturing system



5. The workpiece is carried out by the conveyor CV2.

A conceptual Petri net model is first chosen, which describes the aggregate manufacturing process. The places which represent the subtasks indicated as the task specification are connected by arcs via transitions in the specified order corresponding to the flow of subtasks and a workpiece. The places representing the machines are also added to connect transitions which correspond to the beginning and ending of their subtasks. Thus at the conceptual level the manufacturing process is represented as shown in Figure 9. In this step, if necessary, control conditions such as the capacity of the system between the respective subtasks must be connected to regulate the execution of the Petri net. For the cell with one robot, the place "Robot" has two input transitions and two output transitions, but these transitions are not firable at the same time, so they are not in conflict for firing. The firing of only one of these transitions is permitted using gate arcs (Yasuda, 2008). Next, each place representing a subtask at the conceptual level is translated into a detailed subnet. Figure 10 shows the detailed Petri net representation of loading, processing and unloading in Figure 9.

Figure 9. Petri net representation of the example system at the conceptual level

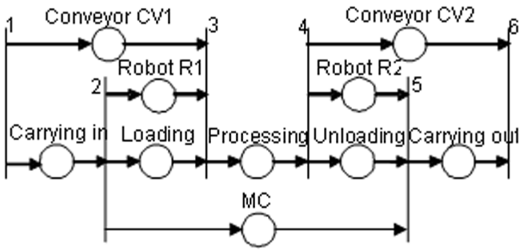
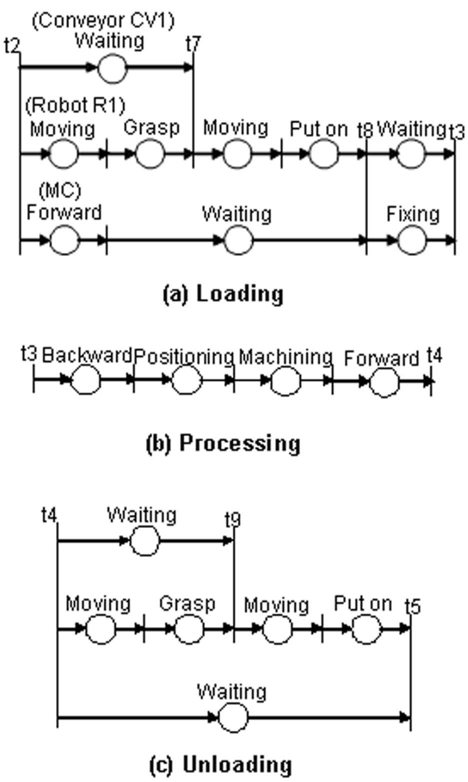


Figure 10. Detailed Petri net representation of subtasks



For the example system, the hierarchical and distributed control system has been realized using a set of PCs. Each machine controller is implemented on a dedicated PC. The station controller is implemented on another PC. Communications among the controllers are performed using serial communication interfaces.

The names of global transitions and their conflict relations are loaded into the coordinator in the station controller. The connection structure of a decomposed Petri net model and conflict relations among local transitions are loaded into the Petri net based controller in the corresponding machine controller. In the connection structure, a transition of a Petri net model is defined using the names of its input places and output places; for example, t1-1=p1-1, -p1-11, where the transition no.1 (t1-1) of Robot controller (subsystem no.1) is connected to the input place no.1 and the output place no.11. For the distributed control system shown in Figure 3, the Petri net representations assigned to the machine controllers are shown in Figure 11. Structural information for the robot controller inputted to the loader is shown in Table 1.

Using the names of transitions in the subsystems, global transitions are defined; for example, G2: t0-2, t1-21, t2-22, t3-23 indicates that the global transition G2 is composed of the transition

Figure 11. Petri net representation of machine controllers (: global transition, : local transition)

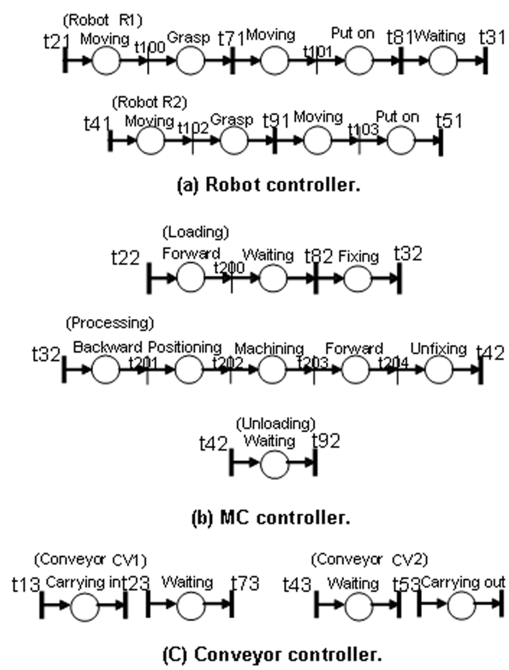


Table 1. Structural information for the robot controller, as inputted to the loader

t1-21=-p1-11	t1-41=-p1-21
t1-100=p1-11, -p1-12	t1-102=p1-21, -p1-22
t1-71=p1-12, -p1-13	t1-91=p1-22, -p1-23
t1-101=p1-13, -p1-14	t1-103=p1-23, -p1-24
t1-81=p1-14, -p1-15	t1-51=p1-24
t1-31=p1-15	

no.2 of Station controller (subsystem no.0), the transition no.21 of Robot controller, the transition no.22 of MC controller (subsystem no.2), and the transition no.23 of Conveyor controller (subsystem no.3). Then, the coordinator information for the example distributed control system is shown in Table 2.

By executing the coordinator and Petri net based controllers algorithms based on loaded information, simulation experiments have been performed. The robot controller executes robot motion control through the transmission of command. The MC controller and the conveyor controller communicate with a dedicated PLC. The Petri net simulator initiates the execution of the subtasks attached to the fired transitions through the serial interface to the robot or other external machine. When a task ends its activity, it informs the simulator to proceed with the next activations by the external permissive gate arc (Yasuda, 2010). The detailed Petri net representation for real-time external machine control is shown in Figure 12. External permissive gate arcs from sensors for detecting the completion of work handling are employed as shown in Figure 13.

Table 2. Coordinator information for the example distributed control system

G1: t0-1, t3-13	start of carrying in
G2: t0-2, t1-21, t2-22, t3-23	start of loading from CV1
G3: t1-71, t3-73	end of grasp on CV1
G4: t1-81, t2-82	end of putting on MC
G5: t0-3, t1-31, t2-32	end of loading into MC
G6: t0-4, t1-41, t2-42, t3-43	start of unloading from MC
G7: t1-91, t2-92	end of grasp on MC
G8: t0-5, t1-51, t3-53	end of putting on CV2
G9: t0-6, t3-63	end of carrying out

The machine controllers control two conveyors or robots, so control software on each PC is written using multithreaded programming. Petri net simulation and task execution program through serial interface are implemented as threads in each machine controller. Real-time task execution using multithreaded programming is shown in Figure 14.

In multithreaded system a thread is an independent flow of execution. All threads share code, data, stack, and system memory areas. So, all threads can access to all global data. Further, because each thread has a separate CPU state and its own stack memory, all local variables and function arguments are private to a specific thread. Context switching between threads involves simply saving the CPU state for the current thread and loading the CPU state for the new thread. It's easy for threads to interact with each other by way of synchronization objects and intertask communications, because they have some shared memory. Memory management data structures do not need to be changed, because the threads share the same memory areas (Grehan, et al. 1998).

In the implementation both the Petri net simulation and task execution threads access to the external gate variables as shared variables; the task execution thread writes the new values of the gate variables and the Petri net simulation thread reads them. Mutual exclusive access control is implemented, so that, while one thread accesses to the shared variables, the other thread can not access to them. Control software using multithreaded programming was written in Visual C# under OS Windows XP SP3 on a general PC. Experiments using a real industrial robot show that the Petri net simulation thread and the task execution threads proceed concurrently with even

Figure 12. Detailed Petri net representation of real-time machine control

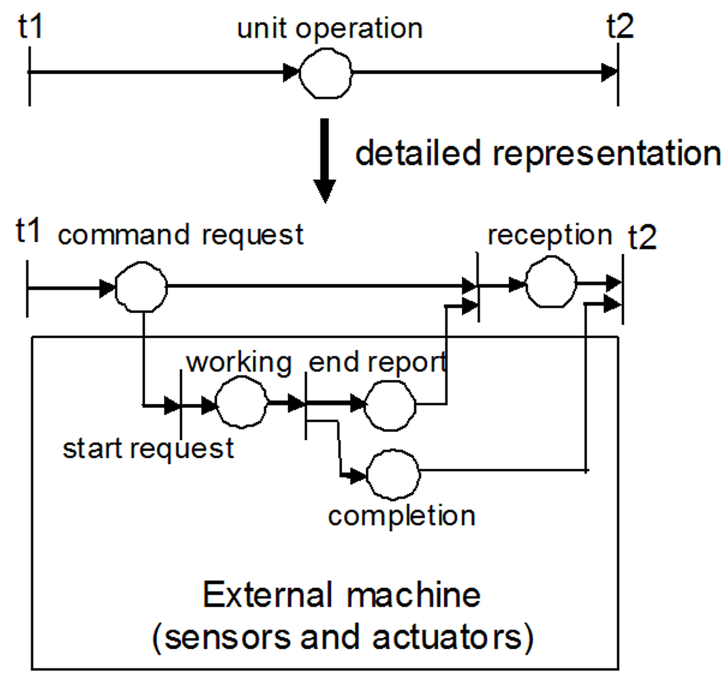


Figure 13. External permissive gate arcs from sensors for work handling

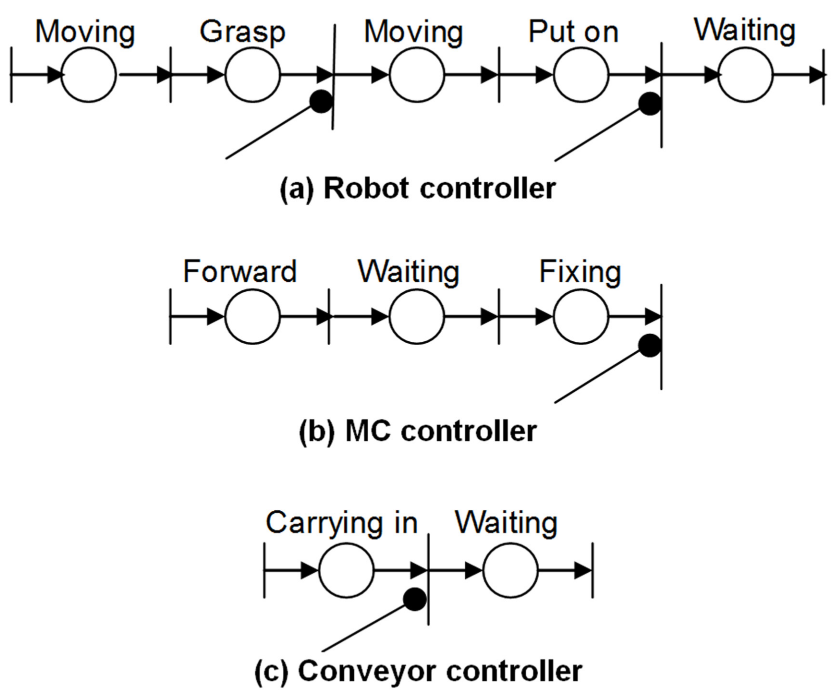
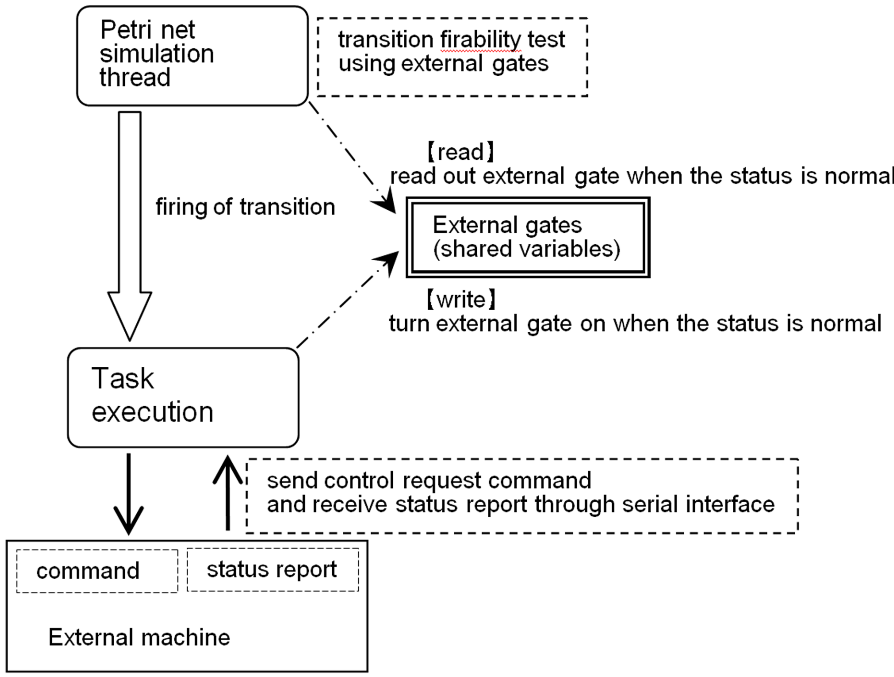


Figure 14. Real-time execution of subtasks using multithreaded programming



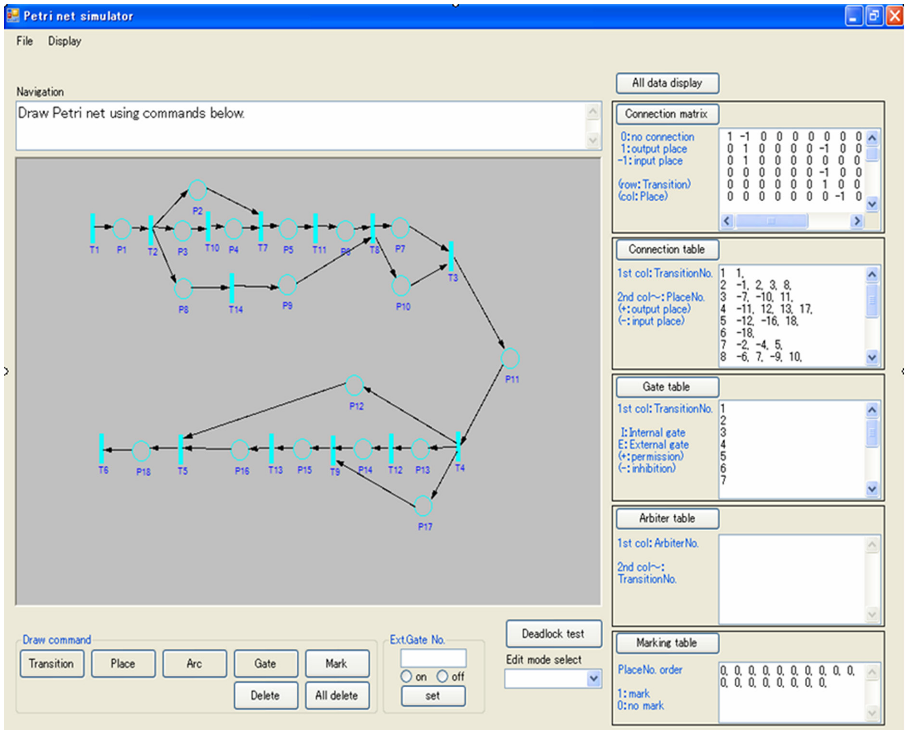
priority, and the values of external gate variables are changed successfully; after the task execution threads write the new values, the Petri net simulator thread reads them immediately.

Experimental results show that the decomposed transitions fire simultaneously as the original transition of the detailed Petri net of the whole system task. The robots cooperated with the conveyors and the machining center, and the example manufacturing cell performed the task specification successfully. Firing transitions and marking of tokens can be directly observed on the display at each time sequence using the simulator on each PC (Yasuda, 2009), as shown in Figure 15. The Petri net modeling thread is executed as the main thread to perform the modeling, drawing and modification of the overall system control net model based on task specification. When the transformation of graphic data of the Petri net model into internal structural data is finished, the Petri net simulation thread starts. During simulation and task execution, liveness or deadlock is possibly decided, and the user can stop the task execution at any time.

CONCLUSION

A methodology to construct hierarchical and distributed control systems, which correspond to the hardware structure of manufacturing cells, has been proposed. By introduction of the coordinator, the Petri net based controllers are arranged according to the hierarchical and distributed nature of the overall manufacturing system; the coordination mechanism is implemented in each layer repeatedly. The Petri net model in each Petri net based machine controller is not so large and easily manageable. The overall control structure of the example robotic manufacturing cell was implemented on a communication network of PCs using multithreaded programming. In accordance with the implementation using multithreaded programming, hierarchical and distributed implementation

Figure 15. View of Petri net simulator



under a real-time operating system on a network of microcomputers connected via a serial bus is also possible, where each microcomputer is dedicated to the local Petri net model of a subsystem in the manufacturing cell. Thus, modeling, simulation and control of large and complex manufacturing systems can be performed consistently using Petri nets.

The proposed methodology has the following advantages especially for manufacturing cells composed of several parallel processes.

1. Because of decomposition of detailed Petri net model, design and implementation of lower level controllers can be performed efficiently for each subsystem. Modification or breakdown in a lower level controller is restricted to the corresponding controller. Machine controllers can be realized using conventional relay circuits, PLCs, or microcomputers. From the practical point of view, transformation Petri nets into ladder diagrams, PLC or microcomputer programs, and inversely transformation them into Petri nets, are desirable.
2. By monitoring the flow of token in the upper level controller, the global state of the system can be conceptually understood.
3. By realizing the upper level controller using a general PLC or computer, the Petri net model (places and transitions) can be easily changed according to task specification.

From the proposed coordinator algorithm the coordination mechanism is performed with the master-slave multicomputer architecture; the upper level controller is the master and the lower level controllers are slaves. So access conflict and lock out do not occur, and both hardware and software structures can be simply realized, especially including non-homogeneous system. Although in the master-slave architecture of operating systems the performance of the upper level controller is required to be sufficiently high, there is no problem because robotic arms, conveyors and machine tools are very slow in comparison with the controller. In case of an accident or breakdown of the upper level controller, manual operation modes must be provided for the lower level controllers. In order to improve the Petri net based control system for manufacturing cells, future works include the following:

1. The number of token is increased and some kinds of tokens are provided in a place.
2. The time required for transition firing condition in a place is provided.

The above extensions of the Petri net is necessary to optimize the process control in a manufacturing cell using simulations without real machines. It is also possible to analyze deadlock phenomena in Petri net models based on detailed simulations.

REFERENCES

Bruno, G., & Marchetto, G. (1986). Process-translatable Petri nets for the rapid prototyping of process control systems. *IEEE Transactions on Software Engineering*, *SE-12*(2), 346–357.

Crockett, D., Desrochers, A., Dicesare, F., & Ward, T. (1987). Implementation of a Petri net controller for a machine workstation. *Proceedings of 1987 IEEE International Conference on Robotics and Automation*, (pp.1861-1867).

Desrochers, A. D., & Al-Jaar, R. Y. (1995). *Applications of Petri Nets in manufacturing systems: Modeling, control and performance analysis*. IEEE Press.

Grehan, R., Moote, R., & Cyliax, I. (1998). *Real-time programming A guide to 32-bit embedded development*. Reading, Massachusetts: Addison-Wesley.

Hasegawa, K., Takahashi, K., Masuda, R., & Ohno, H. (1984). Proposal of mark flow graph for discrete system control. *Trans. of SICE*, *20*(2), 122–129.

Lee, E. J., Togueni, A., & Dangoumau, N. (2006). A Petri net based decentralized synthesis approach for the control of flexible manufacturing systems. *Proceedings of the IMACS Multiconference Computational Engineering in Systems Applications*.

Martinez, J., Alla, H., & Silva, M. (1986). Petri nets for the specification of flexible manufacturing systems. In *Modeling and Design of Flexible Manufacturing Systems* (pp. 389–406). New York: Elsevier Science Publishing.

Miyagi, P. E., Hasegawa, K., & Takahashi, K. (1988). A programming language for discrete event production systems based on production flow schema and mark flow graph. *Trans. of SICE*, *24*(2), 183–190.

Murata, T., Komoda, N., Matsumoto, K., & Haruna, K. (1986). A Petri net based controller for flexible and maintainable sequence control and its application in factory automation. *IEEE Transactions on Industrial Electronics*, *33*(1), 1–8. doi:10.1109/TIE.1986.351700

Reisig, W. (1985). *Petri Nets*. Berlin, Germany: Springer-Verlag.

Silva, M. (1990). Petri nets and flexible manufacturing. In Rozenberg, G. (Ed.), *Advances in Petri Nets LNCS 424* (pp. 374–417). Berlin, Germany: Springer-Verlag.

Silva, M., & Velilla, S. (1982). Programmable logic controllers and Petri nets: A comparative study. In Ferrate, G., & Puente, E. A. (Eds.), *IFAC Software for computer control* (pp. 83–88). Oxford, UK: Pergamon.

Valette, R., Courvoisier, M., Begou, J. M., & Albukerque, J. (1983). Petri net based programmable logic controllers. *Proceedings of 1st International IFIP Conference on Computer Applications in Production and Engineering*, (pp. 103-116).

Yasuda, G. (2008). Hierarchical and distributed control of robotic manufacturing processes based on Petri nets. *Proceedings of the IEEE International Conference on Automation and Logistics*, (pp.221-226).

Yasuda, G. (2009). Design and distributed control of discrete event robotic manufacturing systems using Petri nets. *Proceedings of the 2009 IEEE International Conference on Automation and Logistics*, (pp.1861-1866).

Yasuda, G. (2010). Petri net model based specification and distributed control of robotic manufacturing systems. *Proceedings of International Conference on Advanced Mechatronics 2010* (ICAM 2010), (pp. 410-415).

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 385-400, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 24
Lean Thinking Based Investment Planning at Design Stage of Cellular/Hybrid Manufacturing Systems

M. Bulent Durmusoglu
Istanbul Technical University, Turkey

Goksu Kaya
Istanbul Technical University, Turkey

ABSTRACT

This chapter focuses on providing a methodology for lean thinking based investment planning from the perspective of cellular or hybrid manufacturing systems. The chapter has been divided into three parts. First part provides a general explanation of why lean thinking is so beneficial for managing manufacturing processes and obtaining sustained improvement. This part then moves to the aim of cell formation, and then uses value stream mapping to map current state for visualizing material-information flow and to design a desired future state for examining economic aspects of new machine investment decisions aligned with lean manufacturing principles. The purpose of second part is to explore axiomatic design approach; it provides an overall view of what to do. The third part presents the actual use of the methodology with implementation of hybrid system at a furniture factory; it helps to see application results of this methodology as part of a lean manufacturing program.

INTRODUCTION

Traditional manufacturing systems are built on a functional layout or an assembly line with the principle of economies of scale. This point of view causes much capital investments in high-volume operations and large work-in-process inventories. As an alternative to traditional manufacturing, the principles of the Toyota Production System (TPS) have been widely adopted in recent years. Application of TPS principles have led to lean manufacturing (Sullivan *et al.*, 2002). Womack & Jones (1996) used the term lean thinking as the thinking process of Taiichi Ohno and the set of

DOI: 10.4018/978-1-4666-1945-6.ch024

methods describing the Toyota Production System (Womack & Jones, 1996; Monden, 1993). Lean manufacturing emerged as a global approach that uses different tools to focus on waste elimination and to manufacture products that meet customer's needs (Hines & Taylor, 2000). Lean manufacturing has been increasingly adopted as a potential solution for many organizations, particularly within the automotive industries (Womack, *et al.*, 1990; Day, 1998; Jones, 1999) and aerospace industries (Abbett *et al.*, 1999; Womack & Fitzpatrick, 1999).

Lean production requires the analysis of the "value stream". A value stream is defined as all the value-added and non-value-added operations required manufacturing specific products and services to a customer (Womack & Jones, 1996; Rother & Shook, 1998). Value Stream Map is an enterprise improvement technique to visualize entire production process, representing information and material flows, to improve the production process by identifying waste and its sources (Rother & Shook, 1998).

Lean manufacturing focuses on the waste elimination and produces products that meet customer expectations. Lean production uses production and assembly cells consisting of product focused resources. The aims of the cell formation are smoothing work flow with flexible operation across a wide variety of low cost and high quality products by means of waste elimination. Economic benefits of lean manufacturing include smaller floor space requirements, lower work-in process, reduced lead-times and higher throughput (Sullivan, *et al.*, 2002). Lean production focuses on value pulled from the next upstream activity as customer. As value is specified, value streams are identified eliminating steps that do not create value, so the product will flow smoothly toward the customer. A value stream mapping is an enterprise improvement technique to visualize an entire production process by identifying waste (Braglia *et al.,* 2006).

Cellular manufacturing is an important technique in the planning and controlling of manufacturing system. Cellular manufacturing offers three groups of benefits. These benefits are: human related factors facilitated by empowerment in smaller cells; improved flow and supervisory control in cells to deal with smaller number of parts and facilities; improved operational efficiency, obtainable due to similarity; setup reduction; batch size reduction; improvement in performance related to productivity, quality and agility (Babu *et al.*, 2000). In practice, it is usually hard to partition all machines into independent cells. So, a functional layout generally becomes necessary. Because of this fact, hybrid cellular manufacturing systems (HMS) are required (Suresh 1991).

Hybrid manufacturing system (HMS) is the system where manufacturing cells and functional layout coexist (Shambu & Suresh, 2000), and also it has an advantage of more product flexibility (Satoglu *et al.,* 2009) and less capital investment for machines. Utilization of alternative machines in the HMS reduces additional machine purchasing requirements, and therefore it is beneficial (Satoglu *et al.,* 2009).Empirical evidences also show that hybrid manufacturing system is common for practice (Marsh *et al.,* 1999).

Lean manufacturing tools and techniques provide economical basis to managers for investment planning decisions. Value stream mapping creates a common language about a production process, enabling more purposeful decisions to improve the production system. Value stream map of the system should be taken into account for the design of future state to examine the economic aspects of new machine investment decisions (Sullivan, *et al.*, 2002).This chapter attempts to provide insight as to the choice and use of appropriate tools for designing a successful lean manufacturing system. Although it does not cover every lean manufacturing aspect, it does offer a road map that can guide a company for effective new machine investment decisions toward the development of a lean manufacturing environment.

Investment planning is the determination of suitable machines for manufacturing of part families within the cell. Investment planning for cellular manufacturing can enhance the operating characteristics of the system. After equipment purchases are decided, it is necessary to measure the effectiveness of this decision by evaluating the effects of the new equipment on the system (Gosh, 1989). In existing factories, investment decisions for purchasing new machines in favor of cellular or hybrid manufacturing systems can be troublesome to managers. Many of investment decisions in manufacturing industry are not beneficial and economical (Sullivan, *et al.*, 2002). For this reason, this chapter aims to be a guideline to managers for effective new machine investment decisions in implementation of cellular or hybrid manufacturing layout. This chapter focuses on providing a methodology for lean thinking based investment planning in restructuring manufacturing environment from the perspective of cellular or hybrid manufacturing layout. To do this, we used axiomatic design methodology for creating a systematic perspective.

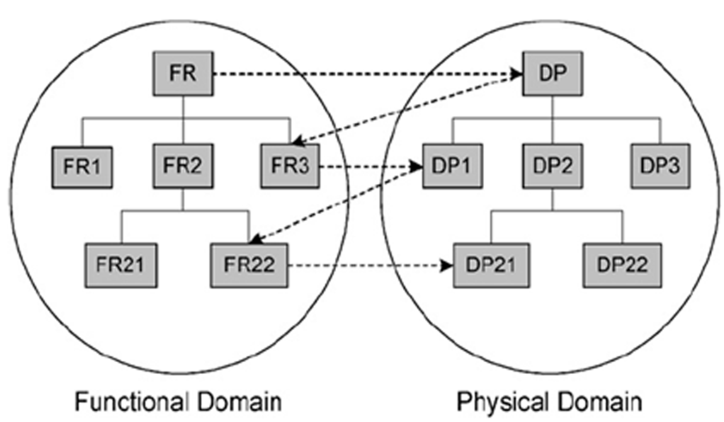
Axiomatic design (AD) is a design theory that was created by Professor Nam Pyo Suh of the Massachusetts Institute of Technology (Suh 1990). The goal of AD is to establish a scientific basis for design and to improve design activities by providing logical and rational processes and tools to designer. In accomplishing this goal, AD provides a systematic search process through the design space to minimize the random search process and determines the best design solution among many alternatives (Kulak *et al.*, 2005).

Some of algorithmic methods can be effective if the design has to satisfy only one functional requirement, but when many functional requirements must be satisfied at the same time, they are less effective. Axioms provide the boundaries within which these algorithms are valid, in addition to providing the general principles. AD includes four domains: customer domain, functional domain, physical domain and process domain. In the customer domain, there are customers' needs or attributes (CAs) from a product, service or system. The functional domain includes customer needs transformed into functional requirements (FRs). To answer FRs, physical domain has design parameters (DPs). Finally, process domain is characterized by the process variables (PVs) to develop a process for production (Suh, 2001). These functional domain and physical domain and mapping between them are illustrated in Figure 1.

The most important concept in axiomatic design is the existence of the design axioms. These axioms are (Suh, 2001):

1. **Axiom:** Independence axiom: The FRs are defined as the minimum set of independent

Figure 1. Hierarchical structure of axiomatic design (Durmusoglu & Kulak, 2008)



requirements that the design must satisfy. A set of FRs is the description of design goals. The independence axiom states that when there are two or more FRs, the design solution must be such that each one of the FRs can be satisfied without affecting the other FRs. That means we have to choose a correct set of DPs to be able to satisfy the FRs and maintain their independence (Suh, 2001).

Axiom: Information axiom: The design that has the smallest information content is the best design. The information content is defined in terms of probability; this axiom also states that the design that has the highest probability of success is the best design (Suh, 2001).

$$\{FR\} = [A]\ \{DP\} \tag{1}$$

Here, {FR} is the functional requirement vector, {DP} is the design parameter vector, and [A] is the design matrix that characterizes the design. The structure of [A] matrix defines the type of design being considered. In order to satisfy the independence axiom, [A] matrix should be an uncoupled (a diagonal matrix) or decoupled design (a triangular matrix) (Suh, 2001).

The design of an ideal manufacturing system depends on the selection of functional requirements (FRs) the system must satisfy within a given set of constraints (Cs). Therefore, an ideal manufacturing system design is not time invariant. It changes with the selection of specific sets of FRs and Cs. An efficient manufacturing system must utilize things, people and information in a rational manner, consistent with basic principles (Suh, 2001).

BACKGROUND

There have been many publications on lean manufacturing and cellular manufacturing systems. Some subsequent steps need for a complete cell formation as Durmusoglu and Nomak (2005) observed in their CMS design implementation. Nancy and Wemmerlöv (2002), published a paper for defining part families and related machine groups, and Sarher and Mondal (1999) studied the evaluation of grouping efficiency measures. Aneke and Carrie (1986) studied on designing of multi product flow lines on the basis of one piece flow.

Many AD applications in designing products, systems, organizations and software have appeared in the literature so far. AD theory and principles have been introduced first time by Suh (1990). Suh (1997) described a conceptual approach for designing of the systems using AD methodology. Suh *et al.* (1998) provided an AD-based model for an ideal production system based on lean principles. Cochran, Eversheim *et al.* (2000) used lean principles to structure smaller production segments by AD. Houshmand & Jamshidnezhad (2002) also provided a lean manufacturing based production system design model using AD approach. In this model, organizational capabilities, technological capabilities and value stream analysis are used as the basis. Kulak *et al.* (2005) published a paper for a complete cellular manufacturing system design methodology based on axiomatic design principles.

Sullivan *et al.* (2002) illustrated an equipment replacement decision problem within the context of lean manufacturing implementation. In particular, they demonstrated how the value stream mapping (VSM) tools can be used to map the current state of a production line and design a desired future state. Further, they provided a roadmap for how VSM can provide necessary information for analysis of equipment replacement decision problems encountered in lean manufacturing implementation.

Considering the literature mentioned above, a road map including a systematic design model for equipment investments in cellular/hybrid manufacturing system is not found. In this study, a methodology is developed using AD principles in order to fill this gap.

LEAN THINKING BASED INVESTMENT PLANNING BY AXIOMATIC DESIGN FOR CMS OR HMS

Main focus of the chapter is giving a systematic methodology by axiomatic design principles to design a desired future state aligned with lean manufacturing principles by using value stream mapping and its associated tools and to provide a roadmap for investment decisions in cellular/hybrid manufacturing.

The first step of axiomatic design is defining the functional requirement at the highest level of the system hierarchy of functional domain (Suh *et al.*, 1998).

Step 1. Defining Functional Requirements (FR1): The design goal of the production system (functional requirement at the highest level) was defined as;

FR1= Improve the system performance

The companies, focusing on lean, aim eliminating wastes and improving system for responding quickly to customer needs.

Step 2. Mapping of FRs in the Physical Domain (DP1): Design parameter (DP), which satisfies the FRs established in the previous step, is selected as below.

DP1= Design of a lean manufacturing system

Step 3. Decomposition of FR (FR1): If the DPs can not be implemented without further clarification, the AD principles recommend returning to the functional domain for decomposing the FRs into their lower functional requirement set (Suh, 2001). The following functional requirements are defined for decomposing the FR determined in the first step.

- FR11= Define customer requirements and expectations
- FR12= Make experts and employees conscious on lean thinking
- FR13= Divide production system into sub-systems for simplification of the system
- FR14= Visualize wastes
- FR15= Apply the proposed plan to eliminate waste in production processes and offices

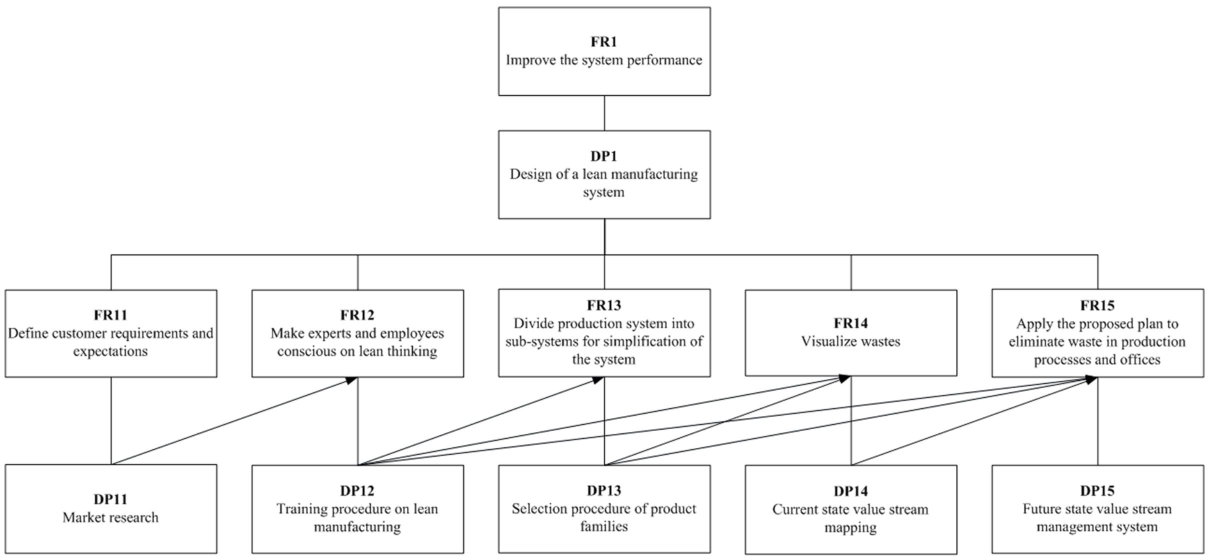
Step 4. Defining the Corresponding DPs of FRs (DP1): We move from the functional domain to the physical domain. The following DPs in response to satisfy the five FR1ns defined in step 3 are listed below.

- DP11= Market research
- DP12= Training procedure on lean manufacturing
- DP13= Selection procedure of product families
- DP14= Current state value stream mapping
- DP15= Future state value stream management system

Step 5. Structuring of Design Matrix (FR1-DP1): The FR-DP sets are defined in Step 3 and Step 4, and the corresponding design matrix (DM) providing the relationship between the FR and DP elements is structured. In the design matrix, a symbol X represents a strong relationship between the corresponding FR-DP pair. It is important to ensure that this DM satisfies the Independence Axiom (IA) of the AD principles. If the DM matrix is uncoupled or decoupled, then it satisfies the Independence Axiom of AD principles (Suh, 2001). The design equation and the DM corresponding to the FR-DP sets are as follows and depicted in Figure 2. Equation 2 is a decoupled design, and satisfies the IA.

$$\begin{bmatrix} FR11 \\ FR12 \\ FR13 \\ FR14 \\ FR15 \end{bmatrix} = \begin{bmatrix} X & 0 & 0 & 0 & 0 \\ X & X & 0 & 0 & 0 \\ 0 & X & X & 0 & 0 \\ 0 & X & X & X & 0 \\ 0 & X & X & X & X \end{bmatrix} * \begin{bmatrix} DP11 \\ DP12 \\ DP13 \\ DP14 \\ DP15 \end{bmatrix} \tag{2}$$

Figure 2. The decomposition FR1-DP1



Defining Customer Requirements and Expectations (FR11-DP11): Making market research is one of the effective working criteria for companies. The companies not giving importance to the research face various risks. Market research is collecting and analyzing of systematic and objective information about products, market and consumers (Megep, 2008). Market research is very important for defining customer demands, sales forecasting in production planning and new product developments.

Making Experts and Employees Conscious on Lean Thinking (FR12-DP12): Employee training is the most efficient method for using human resource effectively in a company. Employee training directly affects company's profitability and decreasing employee turnover by increasing employee satisfaction. Applying of lean thinking and making it company's culture are possible with by training program which makes employees conscious about lean.

Dividing Production System into Sub-systems for Simplification of the System (FR13-DP13): The products should be classified up to the similarities based on production system's characteristics. So, the first step consists of identification of product families and in the selection of one with its production sub-system as the pilot application for improvement. This step continues until all of the product families have been selected.

Pareto Analysis (ABC analysis) is used for separating the vital few from trivial many. Pareto analysis is essential for visualizing volume of the product families in total production. By doing this analysis, it is possible to see how the customer demand is distributed among different product types (Durmusoglu & Kulak 2008). High volume products are responsible for the largest amount of waste's costs (work in process, material handling, other operational costs, etc.). Focusing on these high volume products affects overall performance of the company.

Product family could be a group of high volume products, which pass through the similar operational steps at common machines. Another example of product families is to divide products into catalogued products and specific-project type products. Therefore determining product families depends on production system's characteristics. In order to determine product family, the relationship matrix, including products, production volume, processes, production functions and customers is needed (Durmusoglu & Kulak 2008).

Visualizing Wastes (FR14-DP14): Value stream mapping (VSM) visualizes value and waste resources through the production processes for a defined product family. Therefore, it is used for understanding how the process flow must be, and then, it combines lean tools.

Value stream mapping is a tool, which was created for redesigning production system (Rother & Shook 1998; Womack & Jones, 2002; Pavnaskar *et al.*, 2003). To find causes of waste, it is useful to show parameters for each production processes in detail (Braglia *et al.*, 2006).

Applying the Proposed Plan to Eliminate Waste in Manufacturing Processes and Offices (FR15-DP15): The last step of the proposed methodology is the future state value stream management. For future state, applied for developing the system in redesign stage, the initial aim must be performing a plan that is proposed without any investment. Before the application stage, several scenarios can be tried for deciding new machine investment(s) is needed or not, and finally the cost analysis must be performed based on lead time for decision making. Initial aim of the axiomatic design model is designing cells for products families processed in current machine resources on hand. As seen in the industry, new machine investments to solve bottlenecks of the system are dominated point of view as the key solution. Whereas the majority of new machinery purchasing decisions may create new unnecessary waste resources. This systematic study underlines the necessity of value stream mapping for new investment decisions.

For future state value stream management plan, cell design, kaizen plans, poka-yoke applications, total productive maintenance, heijunka boxes for leveling of production and pull/hybrid system design must be done as lean manufacturing tools with using current resources. Supporting lean manufacturing practices to increase overall performance is possible by lean transition within the whole organization. Therefore, lean office applications have been planned for other units that support production. After the determination of the pilot study group for lean office applications, designing new processes and office cells, and applying visual management practices are aimed to be strengthening with the help of value stream maps and new process flow charts.

Step 6. Decomposition of FR12: FR12 (Make experts and employees conscious on lean thinking) and DP12 (Training procedure on lean manufacturing) are decomposed below:

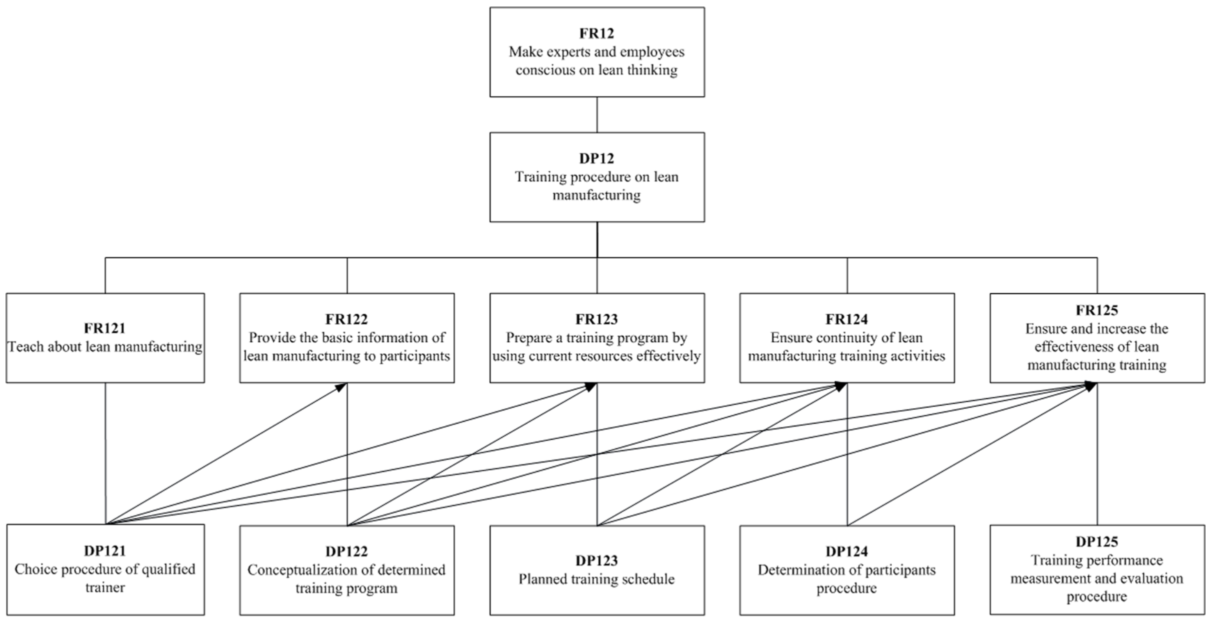
- ◦ FR121= Teach about lean manufacturing
- ◦ FR122= Provide the basic information of lean manufacturing to participants
- ◦ FR123= Prepare a training program by using current resources effectively
- ◦ FR124= Ensure continuity of lean manufacturing training activities
- ◦ FR125= Ensure and increase the effectiveness of lean manufacturing training

Design parameters satisfying the five FRs defined above are as follows;

- DP121= Choice procedure of qualified trainer
- DP122= Conceptualization of determined training program
- DP123= Planned training schedule
- DP124= Determination of participants procedure
- DP125= Training performance measurement and evaluation procedure

The design equation and the DM corresponding to the FR-DP sets are as followed and depicted in Figure 3. This is a decoupled design, and satisfies the IA.

Figure 3. The decomposition FR12-DP12



$$\begin{bmatrix} FR121 \\ FR122 \\ FR123 \\ FR124 \\ FR125 \end{bmatrix} = \begin{bmatrix} X & 0 & 0 & 0 & 0 \\ X & X & 0 & 0 & 0 \\ X & X & X & 0 & 0 \\ X & X & X & X & 0 \\ X & X & X & X & X \end{bmatrix} * \begin{bmatrix} DP121 \\ DP122 \\ DP123 \\ DP124 \\ DP125 \end{bmatrix} \tag{3}$$

Teaching about Lean Manufacturing (FR121-DP121): To increase the efficiency of lean manufacturing and to ensure continuity, employees should be educated on this issue. Planned training program for this purpose is required to be updated continuously to ensure the continuity. Therefore, the experienced instructors from outside or inside the company must be determined for contributing to the preparation and implementation of training programs.

Providing the Basic Information of Lean Manufacturing to Participants (FR122-DP122): Lean manufacturing system, considered to be implemented, brings all the principles of new business culture understanding. To achieve the success of the designed system, employee resistance to new ideas must be eliminated, and then, the new philosophy must be adapted to all employees. Training needs of all employees must be determined and after this, preparation of multi-purpose training programs that meet these needs is necessary.

Preparing a Training Program by Using Current Resources Effectively (FR123-DP123): A timetable for trainers and participants should be identified in parallel with the implementation steps of designed system, and all staff should be informed about lean tools by training programs.

Ensuring Continuity of Lean Manufacturing Training Activities (FR124-DP124): Participants have to be determined according to the company's planned training program and schedule. Selected instructor(s) should determine the training periods and the number of employees to participate in training with a balanced planning in order to ensure the effectiveness of training.

Ensuring and Increasing the Effectiveness of Lean Manufacturing Training (FR125-DP125): At the end of the training, questionnaires should be distributed to evaluate the content of training program. Performances of instructors and training program's content should be reviewed according to the evaluation results obtained from questionnaires. Finally, the scope of further training programs should be revised according to the evaluation results.

Step 7. Decomposition of FR13: FR13 (Divide production system into sub-systems for simplification of the system) and DP13 (Selection procedure of product families) are decomposed as below:

- FR131= Classify products based on customer demand
- FR132= Determine product families based on production system's characteristics

The corresponding DPs satisfying FRs at this step are stated as:

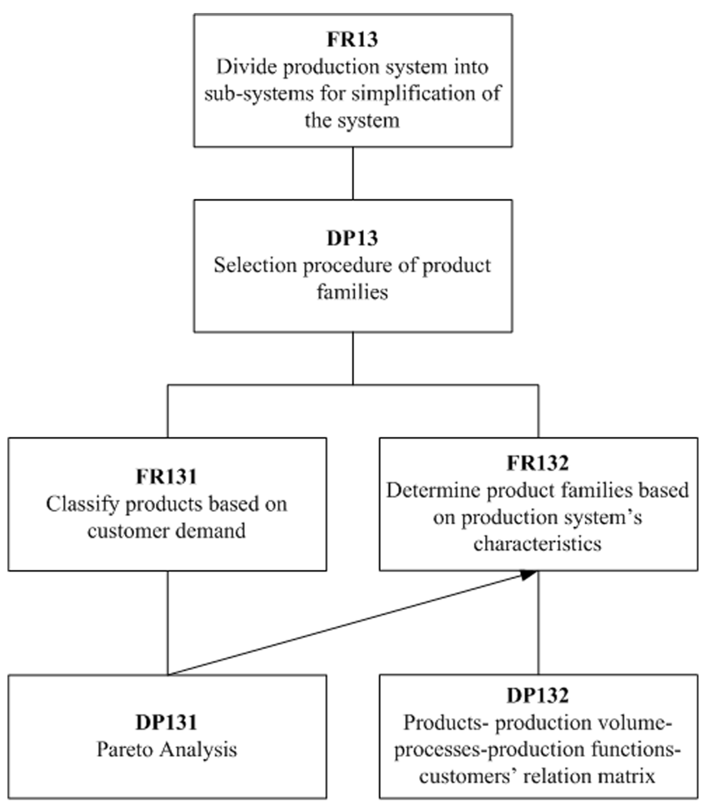
- DP131= Pareto Analysis
- DP132= Products- production volume-processes-production functions-customers' relation matrix

The design matrix for the above set of FRs and DPs are as followed and depicted in Figure 4. This is a decoupled design, and satisfies the IA.

$$\begin{bmatrix} FR131 \\ FR132 \end{bmatrix} = \begin{bmatrix} X & 0 \\ X & X \end{bmatrix} * \begin{bmatrix} DP131 \\ DP132 \end{bmatrix} \quad (4)$$

Classify Products Based on Customer Demand (FR132-DP132): In order to simplify the production system for analysis, products should

Figure 4. The decomposition FR13-DP13



be classified based on customer demand. Pareto (ABC) analysis is used for this classification and it is essential for visualizing volume of the product families. High volume products are responsible for the largest amount of waste's costs (work in process, material handling, other operational costs, etc.). Focusing on these high volume products affects overall performance of the company. So, pareto analysis is used for determining which product families should be selected first for pilot applications.

Determine Product Families Based on Production System's Characteristics (FR131-DP131): Products based on production systems characteristics are grouped as families. For example, products belonging to "A" class, which pass through the similar operations, are grouped as a family. In order to determine product families, constructing a relationship matrix, including products, production volume, processes, production functions and customers, is needed.

Step 8. Decomposition of FR14: FR14 (Visualize wastes) and DP14 (Current state value stream mapping) are decomposed as below:

- FR141= Define process parameters in proposed processes and collect information in response to these parameters
- FR142= Define relations between operations and knowledge processes

The corresponding DP14s satisfying FR14s at this step are stated as followed:

- DP141= Information collecting procedure
- DP142= Value stream mapping procedure

The design equation and the DM corresponding to the FR14-DP14 sets are as followed and depicted in Figure 5. This is a decoupled design, and satisfies the IA.

$$\begin{bmatrix} FR141 \\ FR142 \end{bmatrix} = \begin{bmatrix} X & 0 \\ X & X \end{bmatrix} * \begin{bmatrix} DP141 \\ DP142 \end{bmatrix} \quad (5)$$

Defining Process Parameters in Proposed Processes and Collect Information in Response to These Parameters (FR141-DP141): The first step for the implementation of lean manufacturing tools is visualization of wastes. Before drawing the current state value stream mapping, necessary information such as the process routes, the daily product demands, supplier delivery schedules, available production time, cycle times, setup times, uptime, scrap rates, number of employees, number of shifts, inventory locations and quantities, times between processes, should be collected, and then, should be transferred to value stream maps.

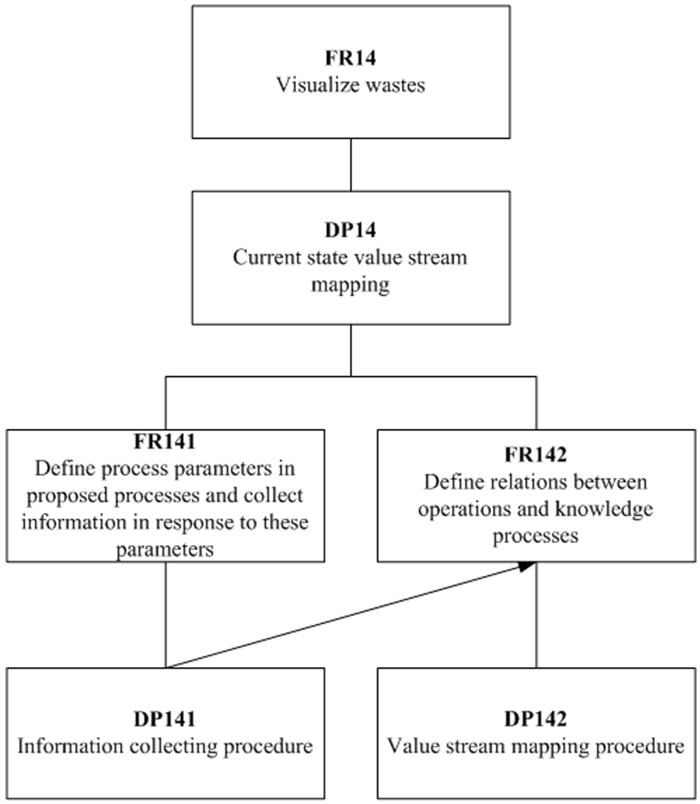
Defining Relations between Operations and Knowledge Processes (FR142-DP142): Drawing current state value stream map helps to visualize the whole system through operation steps by the representation of information and material flow. It also helps to eliminate the non-value added activities. By the use of value stream map, purposeful decisions can be made for the selection of necessary lean tools to eliminate wastes in the system.

Value stream mapping has two parts: big picture mapping and detailed mapping. Before starting detailed mapping of any core process, it is useful to develop an overview of the key features of that entire process. This will (Hines & Taylor, 2000):

- help you to visualize the flows,
- help you to see where waste is,
- pull together the lean thinking principles,
- help you to decide who should be in the implementation teams,
- show relationships between information and physical flows.

Big picture mapping makes easier to understand current state of the system. Big picture mapping consists of five basic steps (Hines & Taylor, 2000):

Figure 5. The decomposition FR14-DP14



Phase 1: Record customer requirements
Phase 2: Add information flows
Phase 3: Add physical flows
Phase 4: Linking physical and information flows
Phase 5: Complete mapping

After big picture mapping, detailed mapping should be done for drawing value stream map to visualize whole system. Collected information should be reflected to the map by value stream mapping symbols.

Step 9. Decomposition of FR15: FR15 (Apply the proposed plan to eliminate waste in manufacturing processes and offices) and DP15 (Future state value stream management system) are decomposed as below:

- FR151= Ensure production pace based on customer demand
- FR152= Ensure the part family flows in product families in the system
- FR153= Ensure one-piece flow
- FR154= Visualize the planned system
- FR155= Increase performance on the systems supporting manufacturing
- FR156= Define other parameters needed for improvement
- FR157= Eliminate the rest trouble bottlenecks
- FR158= Visualize the system with investment
- FR159= Evaluate the system performance with investment
- FR1510= Decrease the excessive volume fluctuations and variations in production
- FR1511= Provide pulling between inter cell flows

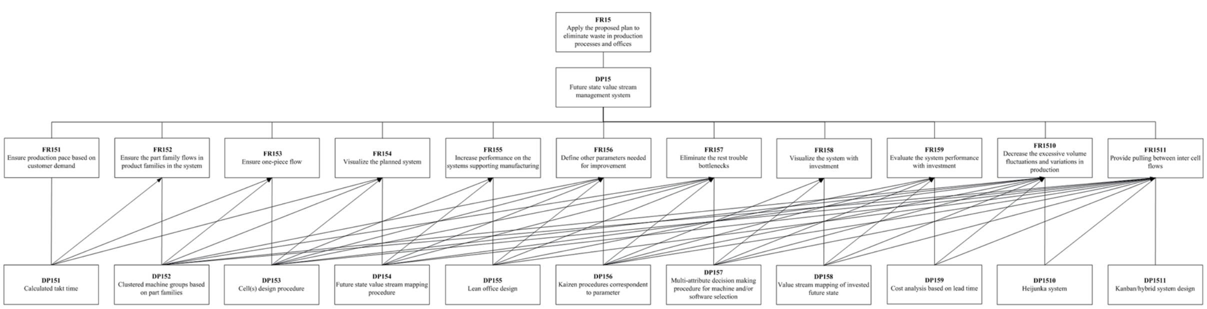
The corresponding DP15s satisfying FR15s at this step are stated as:

- DP151= Calculated takt time
- DP152= Clustered machine groups based on part families
- DP153= Cell(s) design procedure
- DP154= Future state value stream mapping procedure
- DP155= Lean office design
- DP156= Kaizen procedures correspondent to parameter
- DP157= Multi-attribute decision making procedure for machine and/or software selection
- DP158= Value stream mapping of invested future state
- DP159= Cost analysis based on lead time
- DP1510= Heijunka system
- DP1511= Kanban/hybrid system design

The design equation and the DM corresponding to the FR15-DP15 sets are shown in Equation 6 and depicted in Figure 6. This is a decoupled design, and satisfies the IA.

The stage after drawing the current state value stream map is put into practice the lean principles and lean tools by focusing on waste points. Primarily aim at this point will be making planning for elimination of wastes without any investment

Figure 6. The decomposition of FR15-DP15



Equation 6.

$$\begin{bmatrix} FR151 \\ FR152 \\ FR153 \\ FR154 \\ FR155 \\ FR156 \\ FR157 \\ FR158 \\ FR159 \\ FR1510 \\ FR1511 \end{bmatrix} = \begin{bmatrix} X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ X & X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ X & X & X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ X & X & X & X & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & X & X & X & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & X & X & X & X & X & 0 & 0 & 0 & 0 & 0 \\ 0 & X & X & X & X & X & X & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & X & X & X & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X & X & X & X & X & 0 & 0 \\ 0 & X & X & X & 0 & X & X & X & X & X & 0 \\ 0 & X & X & X & X & X & X & X & X & X & X \end{bmatrix} * \begin{bmatrix} DP151 \\ DP152 \\ DP153 \\ DP154 \\ DP155 \\ DP156 \\ DP157 \\ DP158 \\ DP159 \\ DP1510 \\ DP1511 \end{bmatrix}$$

in equipment. At this stage, if equipment investment is needed, the system should be re-planned in accordance to the new situation. The starting point is making a decision to focus on which specific areas (the price reduction request from customers, lead time reduction demand, presence of competing products on the market, product quality problems, etc.). For the purpose of waste elimination plan without any investment, workflow should be simplified, takt time should be calculated to determine the production pace according to customer demand, cell designs should be done, and later other parameters should be taken into account for developing. After configuration plan of the cells, Heijunka and pull system should be established in accordance with the lean principles for reduction of work in process inventories and performing manufacturing just in time according to the customer demand. After this planning stage, the value stream map should be drawn by visualization of the developed system without investment.

After elimination of wastes in production environment, lean office working must be performed in order to increase the performance of systems, supporting manufacturing. By the result of all these activities, if there are still bottlenecks in the system, software or machinery investment decisions were taken into account and the future state value stream map must be revised. Cost analysis based on lead time is recommended for the system's performance evaluation before any investment.

Ensuring Production Pace based on Customer Demand (FR151-DP151): Rather than producing at high-speed, producing according to takt time is preferred (Byrne, 1995). In accordance with the competitiveness conditions, enterprises have to build production systems that will produce the amount that customer demands. Takt time is the cycle time that customers request. Takt time that is calculated by dividing daily operating time to daily customer demand is used to determine the pace of production.

Ensuring the Part Family Flows in Product Families in the System (FR152-DP152): In a manufacturing system for a product family, determination of parts and manufacturing processes/machines assigned to the cell(s) is needed.

For this purpose, there are three approaches:

- Intuitive and Visual Analysis
- Classification & Coding
- Analysis of Routings, Part-Machine Clustering Procedures

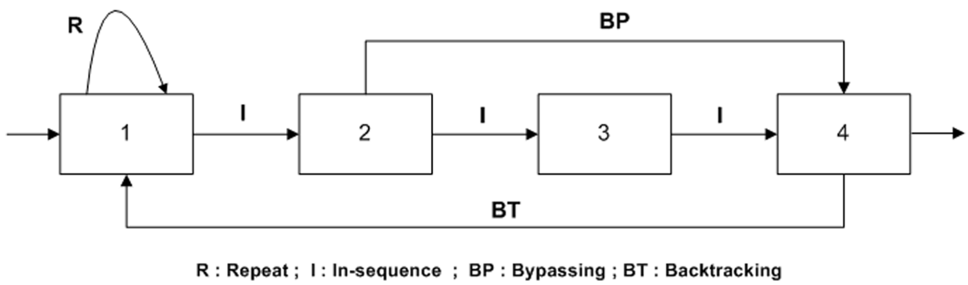
If the system is complex, the clustering procedure is necessary. Using clustering method is more practical and appropriate than alternative classification and coding methods for configurations of cells (Gallagher & Knight 1973).

Ensuring One-Piece Flow (FR153-DP153): Cellular manufacturing system leads to one-piece flow (Miltenburg, 2001). The one-piece flow principle stipulates that parts in a batch travel between machines or processes as single pieces, and do not wait for the rest of the batch to be completed. In other words, part operations on different machines are overlapped and carried out in parallel, which reduces part waiting times, and therefore manufacturing lead-times (Satoglu *et al.*, 2009).

Creating a cellular layout helps to achieve the targeted one-piece flow without repeat and backtracking movements between machines and/or work stations (Satoglu *et al.*, 2009). Figure 7 shows all possible parts movements. If the workflow is complex within the cell, an algorithm, for example, developed by Aneke & Carrie, (1986) would be useful for the design of intra-cell movements.

After using of multi product flow line algorithm developed by Aneke & Carrie, (1986), one piece flow in the cells is ensured. Following the determination of relative positions of the work stations in the cell, the work station layout, preferably U-shaped layout, must be decided. While planning the layout of work stations, transport of materials within the cell and transportation distance should

Figure 7. Four types of product flow (Aneke & Carrie, 1986)



be minimized. Material handling costs in total production costs are ranged from 30% to 75% (Sule,1994). Also, the cells, which have high material transportation traffic between each other, should be placed as close to each other for reducing the number and size of stock area.

Visualizing the Planned System (FR154-DP154): Primary goal is performing the future state plan without any equipment investment. For this purpose, implementation of improvement methods (Kaizen applications, heijunka boxes for production leveling and pull/hybrid system design, reduction of setup times, total productive maintenance, etc.) should be planned. If these improvement methods are not sufficient and there are still some bottlenecks in the production system, new machine investment should be considered. For visualization of the planned system, drawing of future state value stream map is necessary.

Drawing of future state value stream map includes customer demand, workflow and distribution of operations to machines by reducing of work in process inventory, etc.. Moving from the current state value stream map, future state is designed by using lean metrics and tools (Tapping, *et al.*, 2002).

Increasing Performance on the Systems Supporting Manufacturing (FR155-DP155): One of the reasons for long respond time to customer inquiries can be long lead time related to office operations (Suri, 1998). An important method for solving this problem should be lean office. Lean office is elimination of waste from workplace to provide better service to internal and external customers. For reducing non-value added activities and decreasing lead times, the organizational structure should be reconfigured into office cells. At the same time, office cells create a good environment for teamwork. For office cell design, product/service/project families must be determined (with clustering), accordingly the team members should be selected and relevant skill development planning should be done (team building), then the physical office layout should be configured. Moreover, an effective management model should be developed for team work within cells (Durmusoglu & Kulak, 2008).

Defining Other Parameters Needed for Improvement (FR156-DP156): To increase the success in implementation of the planned future state value stream map, Kaizen plans (continuous improvement plans) are recommended (Tapping, *et al.*, 2002). For parameters decided to develop, kaizen tools as 5S for work environment, Single Minute Exchange of Die (SMED) for setup time reduction, Poka-Yokes for reduction of defects, Total Productive Maintenance (TPM) for reduction of potential equipment breakdowns and process kaizens should be applied.

Eliminating the Rest Trouble Bottlenecks (FR157-DP157): After application of kaizen plans, there are still bottlenecks in the system, new investments can be considered. The selection of appropriate machines is very important for prevention of problems caused by production quality, delivery and cost. Many of the decision variables in selection of machines between alternatives make decision making process difficult for managers. In the literature, there are many decision making methods (analytic hierarchy process, second axiom in axiomatic design, etc.) related to machine selection (Saaty, 1990; Babic 1999). On the other side, in the previous step of the axiomatic design model for lean office operations, software investment can be made for accelerating office works, easier information sharing and keeping up to date of office operations.

Visualizing the System with Investment (FR158-DP158): Since the equipment investment decision affects the cell configurations, the future state value stream map should be revised according to the new situation.

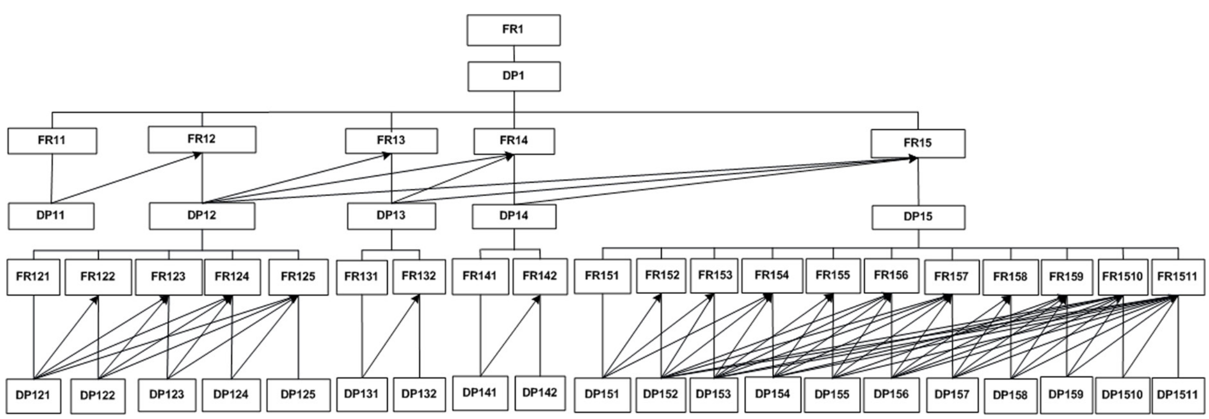
Evaluating the System Performance with Investment (FR159-DP159): Waste elimination in the existing plants and responsibility for new equipment investments are challenging for managers (Sullivan, *et al.*, 2002). Economic analysis based on value stream map should be done in born of the need for equipment investment. Better investment decisions can be made with the use of value stream maps. Cost analysis based on lead time should be made before implementation of future state value stream maps with investment and non-investment decisions. In addition to this, the economic lives of the new equipments affect the investment decision.

Decreasing the Excessive Volume Fluctuations and Variations in Production (FR1510-DP1510): For application of pull production control system according to the lean principles, the production flow must be smoothed primarily. If the flow is leveled according to production-mix throughout the production, the work-in-process inventories and therefore the total lead time will be reduced dramatically. By the accordingly designed Heijunka boxes, the amount of production and how much time it will take to produce this amount can be determined.

Providing Pulling between Inter Cell Flows (FR1511-DP1511): Pull production control system combines the demand and production. Kanban cards are used for the pull production control system. To be successful in Kanban, smoothed production flow is required. Kanban system is a different approach for manufacturing according to the customer demand and communication in the production environment for purchasing the materials needed. Kanban assigned to the cells draws parts from the cells, which produces these parts (Suzaki, 1987).

So far, a road map was presented for lean thinking based investment planning at design stage of cellular/hybrid manufacturing system (see Figure 8). Design matrixes also showed that leaf-level design decisions are consistent.

Figure 8. The decomposition of AD model



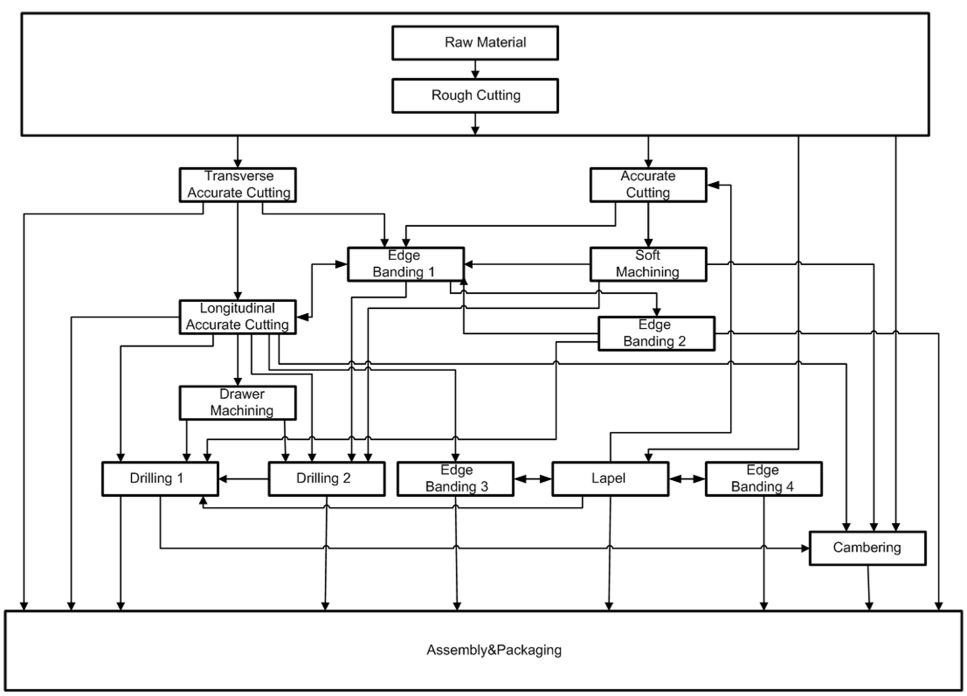
Implementation of the Proposed Methodology

The proposed methodology was tested on a real case application. The proposed methodology was implemented step by step for transforming the existing traditional manufacturing system to cellular/hybrid manufacturing system at a furniture factory. Before implementation, the facility layout of the system was functionally organized (see Figure 9). Plans have been devised for company-wide participation. The first main step (FR11-DP11) was the collection of information about market to define customer requirements. The second main step (FR12-DP12) was making experts and employees conscious on lean thinking by training programs. Around this time, lean manufacturing training programs were prepared and nearly 90% of the employees participated to lectures and workshops about lean tools and techniques.

The third main step (FR13-DP13) was to divide production system into sub-systems for simplification of the system. At this stage, the product families were defined based on manufacturing characteristics. The fourth main step (FR14-DP14) was drawing current state value stream map for the selected product family to visualize wastes and to define which lean tools would be applied to eliminate waste resources. The last main step (FR15-DP15) for the roadmap was applying the proposed plan to eliminate waste in production processes and offices by future state value stream management. The takt time was determined for the selected product family. Determination of parts for that product family and proper allocation of machines for each cell have also been done.

By application of the multi product flow line algorithm developed by Aneke & Carrie (1986), the new layout of the facility was decided based on one piece flow. The new layout consists three newly formed cells and a functional area. Staffing of the cells was finalized regarding of the current cell configurations according to the takt time. Final acquisition in each cell was determined with the

Figure 9. Spaghetti diagram before cell formation



participation of cell teams. Machine locations were defined in the sequence of part movements. After this stage, to increase performance of the systems, supporting manufacturing, lean office studies were planned and started. Before making new investment, the system was tried to improve by kaizen procedures. For maintaining the 5S discipline, 5S evaluation worksheets were designed and routinely implemented. Once the standard works in each cell were defined, proper metrics were developed for continuous improvement monitoring. In order to maintain and improve the performance, equipment effectiveness was calculated and also, total productive maintenance applications and visual management tools were developed. Single minute exchanges of die (SMED) plans were applied to machines to reduce setup times. Beside these, process kaizens were applied continuously for the bottleneck areas.

At this stage, there were still some bottlenecks on the manufacturing system. New machines investment was needed for solving those bottlenecks. Future state value stream map was revised considering new machine investment decision. Finally, four new machines investment is needed to be installed in order to maintain smooth product flow (see Figure 10). Each machine was selected with an AHP based model for the cell and functional area. To evaluate the system performance with investment before purchasing, some economical cost analysis based on lead time was done for the cases of before and after investment. Finally, we decided to make new machine investments by considering the cost analysis results.

At this stage, new machine purchasing procedure was completed, and then, to decrease the excessive volume fluctuations and variations in production, heijunka boxes were planned and to provide continues part flow, a combination of Kanban and push production control system were planned. Based on customer demands and forecasted sales, mixed leveled production plans were established for each cell. Based on these leveled production plans, the convenient batch sizes have been determined for each product family. For Kanban card implementation, new cards were designed. The Kanban cards traffic between the cells and supermarkets were planned. Finally, it has been seen that this methodology is very practical and beneficial for production system perfor-

Figure 10. Spaghetti diagram after cell formation

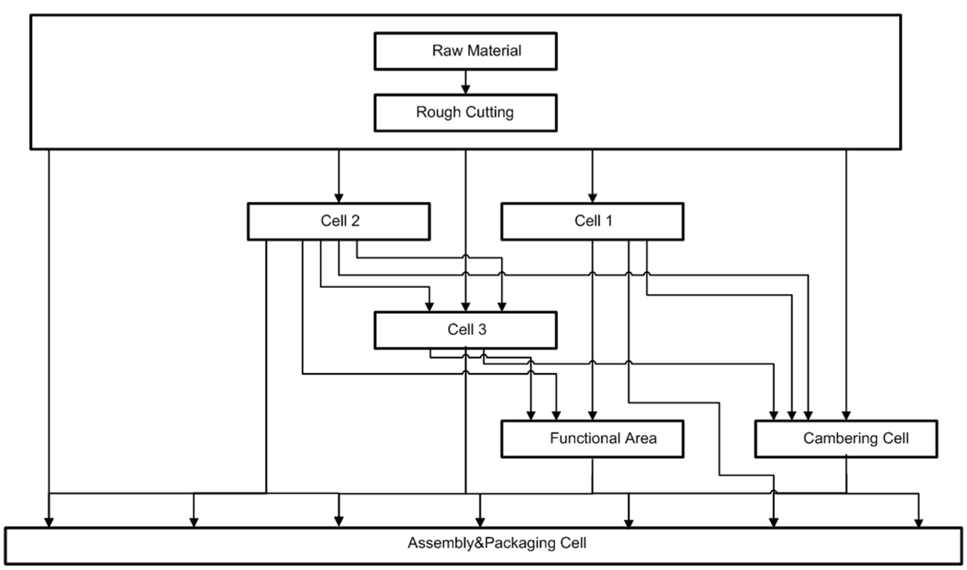


Figure 11. Comparison of business metrics

Performance Criteria	Before Cellular Manufacturing	After Cellular Manufacturing
Lead time (days)	7.5	3.5
Training time (person-hour/year)	7200	16000
Equipment effectiveness (%)	38	85
Number of Poka-Yokes	-	5
Number of Kaizen suggestions (1/year)	4154	4944
Machine breakdown (%)	9.12	5
First time through (%)	50	95
Scrap (%)	3	1
Setup time reduction (%)	0	50

mance improvement and economical investment decisions. The result of this implementation on this real case is listed in Figure 11.

Some of the indicators used in Figure 11 are as explained below:

Manufacturing lead time is called as duration from the moment the chipboards entered to the storage is ready for assembly. It is calculated as following:

$$Lead\,Time = \frac{Work\;In\;\Pr ocess\;Inventory\,(units)}{Demand\;Rate\,(units\;/\;day)} \tag{7}$$

Training time is called as multiplication of the planned duration of training for lean manufacturing applications by the number of participants. Cellular equipment effectiveness (CEE) is the ability to produce high quality and at the right speed at the machineries of cell and to use machines when necessary.

$$Equipment\;Effectiveness = Availability * Performance * Quality \tag{8}$$

$$Availability = \frac{(Total\;Time - Downtime)}{TotalTime} \tag{9}$$

$$Performance\;Efficiency = \frac{Actual\;Run\;Rate}{Ideal\;Run\;Rate} \tag{10}$$

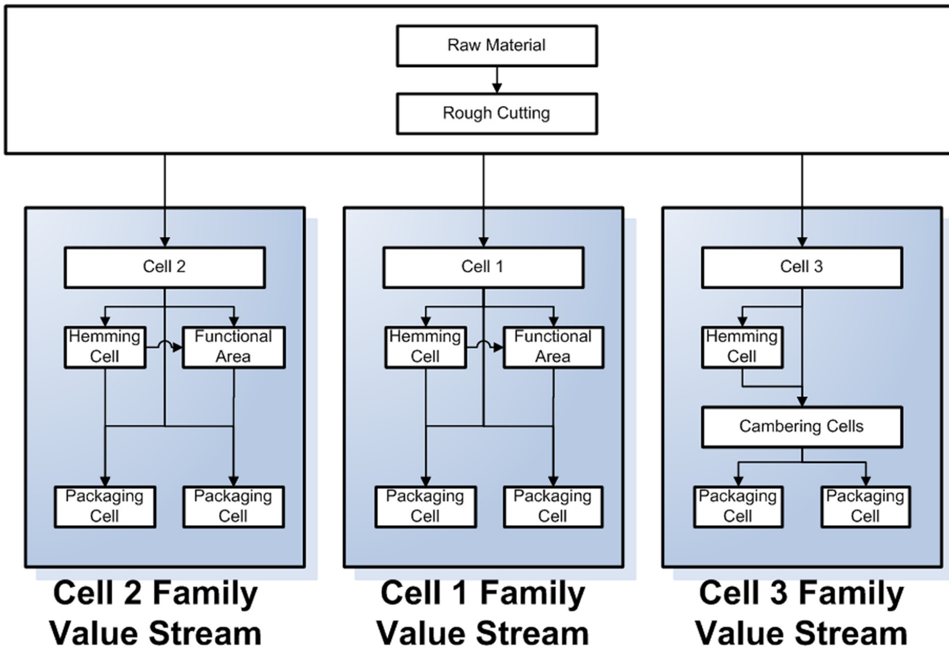
$$Ideal\;Run\;Rate = \frac{1}{Takt\;Time} \tag{11}$$

$$Quality = \frac{Total\,Quantity\,Manufactured - Number\;\;Rejected}{Total\,Quantity\,Manufactured} \tag{12}$$

Number of Poka-Yoke is the number of error preventing applications. Number of Kaizen suggestion is the annual number of suggestions for continuous improvement. First time through is called as the ratio of the number of error-free parts to total quantity manufactured. Scrap rate is the ratio of the number of scrapped parts to total quantity manufactured. Reduction of setup time is the decline in the percentage of duration between the last good part of the previous setup and the first acceptable part in the new setup.

As shown in Figure 11, all the indicators were improved significantly. In order to simplify the work flow further as a continuous improvement effort, a new facility layout has also been planned as shown in Figure 12. This facility layout is expected to decrease the current lead time by 20%. After investigation on the accidents has occurred previously, the number of Poka-Yokes was decided to enhance for the prevention of accidents resulting from misuse of machines. The types of kaizen gifts have been decided to increase. After maturation of lean practices, the number of monthly control tables was planned to increase. For the development related to the considered parameters, it was also decided to form the numerous Kaizen teams.

Figure 12. Spaghetti diagram for future state goal



CONCLUSION

Lean manufacturers must use lean thinking when implementing new purchasing strategies for machines that coincide with the product value stream. Most companies are ignoring the value stream in the decision-making process. In this chapter, we have provided a methodology for transforming a process oriented manufacturing facility into a CMS or HMS with a reasonable investment decision making. The proposed methodology is based on Axiomatic Design principles. Here, a roadmap for how lean thinking based investment planning can be applied at the design stage of cellular/hybrid manufacturing system has been presented. The concept in this chapter can be applied easily to real cases. Value stream mapping is the best tool to visualize wastes and to determine which lean tools can be applied to the system. Visualizing the current state helps to plan the future state by value stream mapping.

In future studies, the proposed methodology for new investment planning can be extended to the process domain which is comprised of the process variables. Decision making when there are multiple axiomatic design alternatives is another research field for the future.

REFERENCES

Abbett, D., Payne, V., & Gulfstream, V. (1999). *Value stream tour*. Atlanta, GA: Lean Enterprise Institute.

Aneke, N. A. G., & Carrie, A. S. (1986). A design technique for the layout of multi-product flow lines. *International Journal of Production Research*, *24*(3), 471–481. doi:10.1080/00207548608919743

Babic, B. (1999). Axiomatic design of flexible manufacturing systems. *International Journal of Production Research*, *37*(5), 1159–1173. doi:10.1080/002075499191454

Babu, A. S., Nandurkar, K. N., & Thomas, A. (2000). Development of virtual cellular manufacturing systems for SMEs. *Logistics Information Management*, *13*(4), 228–242. doi:10.1108/09576050010340866

Braglia, M., Carmignani, G., & Zammori, F. (2006). A new value stream mapping approach for complex production systems. *International Journal of Production Research*, *44*(18–19), 3929–3952. doi:10.1080/00207540600690545

Byrne, A. (1995). How wire mold reinvented itself with kaizen. *Target*, *11*(1), 8–14.

Cochran, D. S., Eversheim, W., Kubin, G., & Sesterhenn, M. L. (2000). The application of ad and lean management principles in the scope of production system segmentation. *International Journal of Production Research*, *38*(6), 1377–1396. doi:10.1080/002075400188906

Day, J. C. (1998). Learning about lean systems at Freudenberg-NOK: Where continuous improvement is a way of life. In Likert, J. K. (Ed.), *Becoming lean* (pp. 179–198). Portland, OR: Productivity, Inc.

Durmusoglu, M. B., & Kulak, O. (2008). A methodology for the design of office cells using axiomatic design principles. *Omega International Journal of Management Science*, *36*, 633–652. doi:10.1016/j.omega.2005.10.007

Durmusoglu, M. B., & Nomak, A. (2005). GT cells design and implementation in a glass mould production system. *Computers & Industrial Engineering*, *48*, 525–536. doi:10.1016/j.cie.2003.03.003

Gallagher, C. C., & Knight, W. A. (1973). *Group technology*. London, UK: Butterworths.

Ghosh, B. K. (1989). Equipment investment decision analysis in cellular manufacturing. *International Journal of Operations & Production Management*, *10*(7).

Hines, P., & Taylor, D. (2000). *Going lean*. UK: Lean Enterprise Research Centre Cardiff Business School.

Houshmand, M., & Jamshidnezhad, B. (2002). Conceptual design of lean production systems through an axiomatic design. *Proceedings of ICAD2002, Second International Conference on Axiomatic Design,* Cambridge, MA (ICAD033).

Jones, D. T. (1999). *Seeing the whole: Macro value stream mapping*. Atlanta, GA: Lean Enterprise Institute.

Kulak, O., Durmusoglu, M. B., & Tufekci, S. (2005). A complete cellular manufacturing system design methodology based on axiomatic design principles. *Computers & Industrial Engineering*, *48*, 765–787. doi:10.1016/j.cie.2004.12.006

Marsh, R. F., Shafer, S. M., & Meredith, J. R. (1999). A comparison of cellular manufacturing research presumptions with practice. *International Journal of Production Research*, *37*(14), 3119–3138. doi:10.1080/002075499190202

Megep. (2008). *Research for marketing*. Project for Improving the Professional Education and Training, Ministry of National Education, Republic of Turkey (In Turkish).

Miltenburg, J. (2001). One-piece flow manufacturing on U-shaped production lines: A tutorial. *IIE Transactions*, *33*(4), 303–321. doi:10.1080/07408170108936831

Monden, Y. (1993). *Toyota production system: An integrated approach to JIT*. Georgia: Engineering Management Press.

Nancy, H., & Wemmerlöv, U. (2002). *Reorganizing the factory* (pp. 95–145). Productivity Press.

Pavnaskar, S. J., Gershenson, J. K., & Jambekar, A. B. (2003). Classification scheme for lean manufacturing tools. *International Journal of Production Research*, *41*, 3075–3090. doi:10.1080/0020754021000049817

Rother, M., & Shook, J. (1998). *Learning to see: Value stream mapping to add value and eliminate muda* (2nd ed.). Brookline, MA: The Lean Enterprise Institute, Inc.

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*, 9–26. doi:10.1016/0377-2217(90)90057-I

Sarher, B. R., & Mondal, S. (1999). Grouping efficiency measures in cellular manufacturing: A survey and critical review. *International Journal of Production Research*, *37*(2), 285–314. doi:10.1080/002075499191779

Satoglu, S. I., Durmusoglu, M. B., & Ertay, T. (2009). A mathematical model and a heuristic approach for design of the hybrid manufacturing systems to facilitate one-piece flow. *International Journal of Production Research*, *48*(17), 5195–5220. doi:10.1080/00207540903089544

Shambu, G., & Suresh, N. C. (2000). Performance of hybrid cellular manufacturing systems: A computer simulation investigation. *European Journal of Operational Research*, *120*(2), 436–458. doi:10.1016/S0377-2217(98)00371-3

Suh, N. P. (1990). *The principles of design*. New York, NY: Oxford University Press.

Suh, N. P. (1997). Design of systems. *Annuals of the CIRP*, *46*(1), 75–80. doi:10.1016/S0007-8506(07)60779-3

Suh, N. P. (2001). *Axiomatic design: Advances and applications*. New York, NY: Oxford University Press.

Suh, N. P., Cochran, D. S., & Paulo, C. L. (1998). Manufacturing system design. *Annuals of the CIRP*, *47*(2), 627–639. doi:10.1016/S0007-8506(07)63245-4

Sule, D. R. (1994). *Manufacturing facilities: Locations, planning and design*. Boston, MA: PWS Publishing Company.

Sullivan, W. G., McDonald, T. N., & Van Aken, E. M. (2002). Equipment replacement decisions and lean manufacturing. *Robotics and Computer-integrated Manufacturing*, *18*, 255–265. doi:10.1016/S0736-5845(02)00016-9

Suresh, N. C. (1991). Partitioning work centers for group technology: Insights from an analytical model. *Decision Sciences*, *22*(4), 772–791. doi:10.1111/j.1540-5915.1991.tb00364.x

Suri, R. (1998). *Quick response manufacturing* (pp. 303–333). Portland, OR: Productivity Press.

Suzaki, K. (1987). *The new manufacturing challenge: Techniques for continuous improvement*. New York, NY: The Free Press.

Tapping, D., Luyster, T., & Shuker, T. (2002). *Value stream mapping*. Portland, OR: Productivity Press.

Womack, J. P., & Fitzpatrick, D. (1999). *Lean thinking for aerospace: the industry that can afford its future*. Atlanta, GA: Lean Enterprise Institute.

Womack, J. P., Jones, D., & Roos, D. (1990). *The machine that changed the world*. New York, NY: Rawson Associates.

Womack, J. P., & Jones, D. T. (1996). *Lean thinking: Banish waste and create wealth in your corporation*. New York, NY: Simon & Schuster.

Womack, J. P., & Jones, D. T. (2002). *Seeing the whole. Mapping the extended value stream*. Massachusetts Lean Enterprise Institute.

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 342-365, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 25
Internal Supply Chain Integration:
Effective Integration Strategies in the Global Context

Virpi Turkulainen
Aalto University, Finland

ABSTRACT

Supply chain relationships and subsequent supply chain integration have received significant attention among both academics and practitioners over several decades. The majority of the prior research has focused on assessing the relationships between the focal manufacturing firm and its suppliers and/or customers. Recent research, however, suggests that successful management of external relationships requires effective integration of the internal supply chain within the focal firm. In this chapter, we focus on internal supply chain integration. The author starts with the assumption that integration is an investment and that the importance of it is context-dependent. By examining data collected in 210 plants in eight countries, the author exploratively examine internal supply chain integration strategies; using cluster analysis, they classify the plants into four groups based on their use of various individual integration mechanisms. In addition, the author assess the use of the four integration strategies under various task contexts and across countries.

INTRODUCTION

Due to recent developments in globalization, companies can be viewed as complex structural networks, which span across different geographical locations and number of businesses and partners. In order to successfully manage those networks, complex managerial decision making structures are needed; in today's globalized environment, companies compete with their capability to integrate knowledge and information, both internally within the organization and externally with their suppliers, customers, and other stakeholders

DOI: 10.4018/978-1-4666-1945-6.ch025

(Galbraith, 2002). The results of prior research also suggest that problems companies are facing in the global economy are often not due to inappropriate strategic analysis but, in fact, are due to failure in designing the organizational structures and subsequent integration of internal and external parties (Bartlett & Ghoshal, 1989; Galbraith, 2000).

Integration of the supply chain has received significant attention among academic researchers and managers since the 1980s (Miles & Snow, 2007). Especially during the recent years, research addressing the management of supply chain relationships has been numerous (e.g., Bozarth, Warsing, Flynn & Flynn, 2009; Cousins & Menguc, 2006; Das, Narasimhan & Talluri, 2006; Germain, Claycomb & Dröge, 2008; Koufteros, Cheng, & Lai, 2007; Vachon & Klassen, 2008; Van der Vaart & Van Donk, 2007). Supply chain integration can broadly be divided into two parts (e.g., Chen & Paulraj, 2004; Fawcett & Magnan, 2002): management of the relationship between a firm and its suppliers and/or customers (external integration) and management of the cross-functional relationship within the manufacturing firm (internal integration).

In this chapter, we focus on internal supply chain integration in manufacturing organizations. The aim of the chapter is to set the stage for effective and efficient management of global supply chain relationships and integration in the global supply network. Integration is often assumed to provide wide benefits and is considered critical in order to survive in the global, constantly changing environment (e.g., Fine, 1998; Koufteros, Vonderembse & Jayaram, 2005). In this chapter, we present a somewhat contrasting view of integration. We build on the idea that integration can be considered as an investment requiring both financial and managerial resources (Adler, 1995; Lawrence & Lorsch, 1967b; O'Leary-Kelly & Flores, 2002). Hence, it is of fundamental importance to understand when integration is particularly relevant and how it can be effectively managed in different contexts. We are especially interested in configurations of individual integration mechanisms, which we label as integration strategies. In this chapter, we address questions such as: How do companies manage the internal integration challenge? Is it possible to identify different integration strategies? Does the use of different strategies differ, depending on the characteristics of the organizational task? Does the use of integration strategies differ across countries? We approach these questions exploratively and analyze empirical data collected in 210 manufacturing organizations in eight countries across the world (Austria, Finland, Germany, Italy, Japan, Korea, Sweden, US).

LITERATURE REVIEW AND THEORETICAL BACKGROUND

Defining Supply Chain Integration

The idea of supply chain integration is not new; integrated supply chain relationships have long been encouraged among manufacturing firms and their supply chain partners (e.g., Lambert, Robeson, & Stock, 1978) and supply chain integration has received significant attention among academic researchers since the 1980s (Miles & Snow, 2007). Despite the long history of research on supply chain relationships and supply chain integration, the perceptions and definitions of it vary widely among scholars. The predominant approach to supply chain management has been to focus on the relationships between a focal firm and its partners (suppliers and/or customers) (Chen & Paulraj, 2004). Yet, others relate supply chain management to the management of both cross-functional relationships within the firm itself and relationships with external parties (e.g., Ballou, Gilbert, & Mukherjee 2000; Barki & Pinsonneault, 2005; Germain & Iyer, 2006; Pagell, 2004; Rodriques, Stank, & Lynch, 2004).

Building on prior definitions of supply chain management (e.g., Chen & Paulraj, 2004), we define supply chain integration as the degree to which a manufacturer collaborates with its supply chain partners (both upstream and downstream) and collaboratively manages *intra*- and *inter*-organizational processes. We make a clear distinction between two aspects of supply chain integration: (1) internal supply chain integration, which refers to the management of intra-organizational relationships and (2) external supply chain integration, which refers to the management of inter-organizational relationships. The goal of supply chain integration, then, is to achieve efficient and effective flows of for example products, services, and information, as well as to manage organizational decision making (Frohlich & Westbrook, 2001).

Prior Research on Supply Chain Integration

Prior research on supply chain integration has focused primarily on external integration. Most of this research has looked at the relationship between a manufacturing firm and its suppliers, taking the perspective of the buyer (Clark, 1989; Cousins & Menguc, 2006; Fine, 2000; Koufteros et al., 2007; Spina, Verganti & Zotteri, 2002). Some have also taken the supplier perspective and addressed the relationship between a manufacturing firm and its customers (Bagchi & Skjoett-Larsen, 2002; Bajaj, Kekre, & Srinivasan, 2004). Others have combined these two approaches and taken a broader perspective, assessing the relationship of a manufacturer to both its suppliers and customers (Rodriques et al., 2004; Swink, Narasimhan & Wang, 2007; Vickery, Jayaram, Droge & Calantone, 2003). Finally, some have assessed the internal supply chain and intra-organizational relationship within the manufacturing firm (Ellinger, Daugherty & Keller, 2000; Kahn & Mentzer, 1996; Pagell, 2004) or both internal and external supply chain relationships either separately (Ballou et al., 2000; Das et al., 2006; Germain & Iyer, 2006; Gimenez & Ventura, 2005; Narasimhan & Kim, 2002) or as an overall measure of supply chain integration (Rosenzweig, Roth, & Dean 2003; Vickery et al., 2003).

In terms of research approach, prior research on supply chain integration can be characterized as looking at the effects of various integration practices on different dimensions of performance. For example, Swink et al. (2007) studied the effects of both customer and supplier integration, conceptualized as the use of various integration mechanisms, on business performance and manufacturing competitive capabilities, and Vachon and Klassen (2008) studied the effects of supplier and customer collaboration on various operational performance measures. Vickery et al. (2003) assessed the effect of supply chain integration on customer satisfaction and financial performance. Several researchers (e.g., Germain & Iyer, 2006; Gimenez & Ventura, 2005; Rodriques et al., 2004) have also related supply chain integration to logistical performance.

Prior research tends to build on the idea that integration has positive performance effects and more emphasis on integration always leads to higher performance (for an excellent exception, see Germain et al., 2008). In the context of internal supply chain integration, Pagell (2004) suggests that "the importance of integration is not in doubt." Even though it has been suggested that the importance of customer, supplier, and internal integration varies from context to context (Das et al., 2006), less attention has been paid to understanding when and how to integrate the supply chain. The benefits of supply chain integration are widely recognized, and integration has been suggested as a strategy to overcome, for example, different thought worlds by incorporating information and inputs from other parties to internal planning (Frohlich & Westbrook, 2001; Swink et al., 2007). However, prior research on the effects of supply chain integration on performance are still somewhat inconclusive and inconsistent (cf.

Das et al., 2006; Germain & Iyer, 2006; Gimenez & Ventura, 2005). This calls for a more detailed analysis of supply chain integration.

Practical evidence from industry suggests that effective internal integration is the first step towards external integration or facilitates it (Gimenez & Ventura, 2005; Stevens, 1989). Evidence also suggests that managers view internal integration as the crux of supply chain management (Fawcett & Magnan, 2002). Furthermore, partly as a response to the inconclusive results of prior research on the performance effects of supply chain integration, Flynn et al. (In press) suggest that successful management of the relationships with suppliers and customers requires that the manufacturing firm is internally integrated; "internal integration provides a vital link between customer and supplier integration, without which companies are unable to reap the full benefits of their SCI [supply chain integration] efforts". This suggests that, in order to benefit from integration with customers and suppliers, firms need to effectively manage their own strategies, practices, and processes in a collaborative manner. Since the prior research has mainly focused on external supply chain integration, the underlying (implicit) idea has been that, despite the firm's engagement in internal integration, external integration provides performance benefits. As an example, let us consider a situation in which a manufacturer integrates processes with a retailer through information systems, such as checkout counter POS system to reduce supply chain costs and uncertainty. However, if the manufacturer is not internally integrated, the advantages of customer integration would not be realized, as the data could not be used across the functional borders of the manufacturer, such as production and inventory control (Germain & Iyer, 2006).

In this chapter, we build on the idea that effective internal integration is indeed needed in order to engage in successful external integration in the supply chain context. We focus on addressing how manufacturing firms can manage the internal integration challenge. Hence, this chapter serves as a basis for increasing understanding of how to successfully manage integration with customers and suppliers. Instead of focusing on the effects of internal integration on performance, however, we are specifically interested in how firms manage internal integration. We call this integration strategy and define it as a configuration of individual integration practices that the manufacturing firm uses.

Few authors have addressed configurations in the supply chain integration context. Frohlich and Westbrook (2001) developed a taxonomy of supply chain integration, "the arcs of integration", based on supplier and customer integration patterns. The taxonomy they presented includes five different types: inward-, periphery-, supplier-, customer-, and outward-facing integration strategies. The taxonomy was developed based on the quartiles of the extent of integration emphasized towards suppliers or customers, excluding the internal aspect of integration. Flynn et al. (in press), on the other hand, present a configuration of supply chain integration based on the general emphasis the firm has put on the three different dimensions of internal, customer, and supplier integration. In this chapter, we approach configurations from a somewhat different perspective; rather than focusing on the overall emphasis on integration, our aim is to understand what managerial mechanisms are used in combination to manage the integration challenge. As stated before, we focus on the internal integration challenge.

Contingency Theory Perspective to Integration

In this research, we build on the contingency theory of organization design (Lawrence & Lorsch, 1967b; March & Simon, 1958; Thompson, 1967) and information processing perspective (Galbraith, 1973; Tushman & Nadler, 1978). In its very basic form, the task-dependent contingency theory suggests that the characteristics of the task determine the most effective organization design and integration (Burns & Stalker, 1961; Child, 1975; Lawrence & Lorsch, 1967a, 1967b; Van de Ven & Delbecq, 1974). Organizational tasks are seen to vary in terms of uncertainty, which is suggested to create varying needs for information processing.

What are the relevant task characteristics? At the overall level, "task" refers to the goal of the organization (Dill, 1958) and it has been characterized for example by uncertainty, predictability, complexity, routineity, analyzability, controllability, and variety (e.g., Drazin & Van de Ven, 1985; Gerwin, 1981; Lawrence, 1981; Perrow, 1967; Tushman & Nadler, 1978; Van de Ven & Delbecq, 1974; Van de Ven, Delbecq, & Koenig, 1976). Based on a literature review, it is possible to trace the overall concept as being uncertainty, which has two sources: unpredictability and complexity (this also resembles the conceptualization made by Bensaou and Venkatraman (1995) in the supply chain context). ***Unpredictability refers to the*** predictability of the task and it can further be divided into task variability and task difficulty (Drazin & Van de Ven, 1985; Van de Ven & Delbecq, 1974; Van de Ven et al., 1976; Woodward, 1965). Task variability is defined as the number of exceptions, whereas task difficulty is defined as the unanalyzability of the search process. Also complexity can further be divided into two dimensions (Lawrence, 1981; Rivkin, 2001): the number of variables in decision making and their interdependence (in order of increasing complexity: pooled, sequential, reciprocal, and team interdependence (Thompson, 1967; Van de Ven et al., 1976)). Recently, Bozarth et al. (2009) conceptualized complexity in the supply chain context as detail and dynamic complexity. Detail complexity can be considered synonymous to the number of elements, whereas dynamic complexity is more related to unpredictability.

According to the information processing perspective, in effective organizations there is a fit between the information processing requirements created by the task characteristics and the capacity of the organization to process information (Galbraith, 1973). ***In order to facilitate information processing, organizations implement vertical and horizontal integration mechanisms***. Integration mechanisms vary in their capacity to facilitate information processing and the richness of the information they facilitate (Daft & Lengel, 1986; Galbraith, 1973). Furthermore, integration mechanisms also vary in their complexity and costs of implementation and use (Galbraith, 1973, 1994; Lawrence & Lorsch, 1967b; Porter, 1985). Integration mechanisms can be classified into five broad categories (Galbraith, 1973; Martinez & Jarillo, 1989) as presented in Table 1. These are presented in order of increasing capacity to facilitate information processing and increasing cost of implementation and use.

Table 1. Classification of integration mechanisms

Integration mechanisms		References
1	Vertical mechanisms of centralization, standardization and formalization	(Burns & Stalker, 1961; Child, 1972; 1973a; 1973b, 1975; Pierce & Delbecq, 1977; Pugh, Hickson, Hinings, & Turner, 1968; 1969)
2	Information systems	(Galbraith, 1973; 1977; 1994)
3	Informal lateral mechanisms, such as cross-functional job rotation	(Edström & Galbraith, 1977)
4	Formal lateral mechanisms, such as cross-functional teams, committees, and integrator roles	(Adler, 1995; Galbraith, 1973; 1977; 1994; Hage, Aiken, & Marrett, 1971; Lawrence & Lorsch, 1967b; Tushman, 1977)
5	Organization-level incentives and other social mechanisms	(Barnard, 1938; Grandori & Soda, 1995)

Instead of focusing on the use of individual integration mechanisms, we approach contingency theory from a configuration perspective (Ketchen, Thomas & Snow, 1993; Meyer, Tsui, & Hinings, 1993; Snow, Miles, & Miles, 2006). Rather than looking at the use of various individual integration tools to manage the internal supply chain integration challenge, the configuration approach assesses the use of integration mechanisms in combination and simultaneously. Although the configurational approach has received less attention in organization research, it is suggested to overcome the critique of contingency theory being a reductionist theory and is seen as a holistic approach (Meyer et al., 1993).

Building on contingency theory and the information processing perspective, it is expected that the use of integration strategies will vary depending on the task characteristics presented above; strategies relying more heavily on the use of complex mechanisms are more likely found in organizations having more uncertain tasks. Furthermore, in order to develop successful global supply chain integration strategies, it is interesting to also understand whether internal supply chain integration strategies differ across countries. However, as we do not know the different strategies and there are no theoretical arguments on what kind of strategies could be found, we cannot present any precise hypotheses about the effect of task characteristics and country variables on the use of integration strategies, but rather must assess the issues empirically and exploratively.

INTERNAL INTEGRATION STRATEGIES

Data and Measures

In order to identify the internal integration strategies implemented by manufacturing firms, we use data collected in the third round of the High Performance Manufacturing (HPM) research initiative (previously known as the World Class Manufacturing Project (Schroeder & Flynn, 2001)). Data were collected from 236 plants in eight countries (Austria, Finland, Germany, Italy, Japan, Korea, Sweden, and the United States) and three industries (electronics, machinery, and transportation) between the years 2003 and 2006. Due to missing data, 210 plants are analyzed in this research. Because of the focus of the research on *internal* integration within the manufacturing firms, the unit of observation is a manufacturing plant. In order to obtain a similar number of plants for each combination of industry and country, stratified sampling (Forza, 2002) was used. The sample characteristics are presented in Table 2. Throughout it duration, the focus of the HPM project has been on world-class manufacturing practices and, hence, the sampling focused on mid-sized or larger plants with more than 100 employees.

The response rate varied across countries, average being approximately 65%. In order to achieve such a high response rate, plant managers were contacted in advance and asked to participate in the survey. In addition, each plant was promised a profile of responses. In these profiles the focal

Table 2. Stratification of the sample

	Country								
Industry	**AUT**	**FIN**	**GER**	**ITA**	**JPN**	**KOR**	**SWE**	**USA**	**Sum**
Electronics	10	14	8	10	9	8	7	9	75
Machinery	5	6	13	9	8	8	10	7	66
Transportation	4	10	19	7	10	7	5	7	69
Sum	19	30	40	26	27	23	22	23	210

organization was compared to the rest of the sample in terms of for example strategy, manufacturing practices, and performance. Data were collected using written survey questionnaires in the native language in each country. Translation and back-translation were carried out to check for consistency across the countries. The data collection instrument was pilot tested with both academics and managers in different countries. The data were collected using multiple respondents per plant; the questionnaire was differentiated to focus on the specific expertise area of each respondent (i.e. key informant method (Bagozzi, Yi, & Phillips, 1991)). In total, 12 different questionnaires were developed. These were distributed to 12-23 individuals in various organizational units and at different levels in the vertical organization (managers, supervisors, and direct labor).

Measurements

Five integration mechanisms were selected for analysis, in this research: centralization of decision making, information systems, cross-functional job rotation, cross-functional teams, and plant-level incentives. These mechanisms represent the five different categories presented in Table 1 and subsequently differ in their capacity to facilitate information processing, as well as in terms of the costs of implementation and use. In this research, the adoption of a particular integration mechanism is assessed as the intensity of use rather than whether a particular mechanism is adopted or not (e.g., Ettlie, 1995; Ettlie & Reza, 1992), because it is emphasized that the level of adoption of a mechanism, such as information systems, can vary. Previous OM literature on integration has also included items referring to the use of several mechanisms, such as job rotation and cross-functional teams, in the same construct (e.g., Swink, 2000; Swink et al., 2007). In this research, different mechanisms are operationalized as separate constructs because it is perceived that different mechanisms can be utilized to a varying extent, independently of each other. This is also required in order to identify integration strategies based on the use of various individual integration mechanisms, in contrast to an overall emphasis on integration mechanisms.

The five integration mechanisms were operationalized using seven-point Likert scales (anchor points as 1 = Strongly disagree, 7 = Strongly agree), except for one (information systems). The respondents were asked to rate specific statements regarding the use of integration mechanisms. Mainly multi-item scales were used. The operationalizations are presented in Table 4 and Table 5.

This research incorporates five measures of task characteristics, which are directly related to the dimensions presented above: task variability, task difficulty, number of variables in decision making, and reciprocal task interdependence. The operationalization of the task characteristics in the context of this research is based on prior research. Links to previous research is presented in Table 3. Two of the measures are objective (introduction of new products, reciprocal interdependence). Three constructs are measured on a 7-point Likert scale (1 = Strongly disagree,…, 7 = Strongly agree), including both single-item (customization of orders) and multi-item scales (modularity of products, introduction of new technology). In case of a subjective single-item measure, multiple respondents were used to avoid single-informant bias. Operationalization of task characteristics are presented in Table 4 and Table 5.

For multi-item scales, each plant-level item value was first calculated as an arithmetic average of multiple (three) responses. Scales were then developed with confirmatory factor analysis (CFA) using structural equation modeling (SEM). Appropriateness of factor analysis is established by high and statistically significant inter-item correlations ($p < 0.01$), high value of the measure of sampling adequacy (MSA), and statistically significant Bartlett test score ($p < 0.01$). Maximum Likelihood estimation was used for SEM. The standardized factor loadings of CFA are pre-

Table 3. Characteristics of tasks and their application to OM context

Dimension	Application to OM context	Explanation	References
Task variability	Customization of orders	Customization causes exceptions in the manufacturing task because the output changes in an unexpected manner; when the degree of customization is high, the manufacturing task is unpredictable and cannot be preplanned requiring information processing across units.	(Forza & Salvador, 2002; Perrow, 1967; Wheelwright & Clark, 1992; Vickery, Dröge, & Germain, 1999)
Task difficulty	Product modularity	Modularity affects the difficulty in performing work when exceptions (customization) occur; when modularity is high, changes in products can be made quickly requiring less information processing across units for example regarding designs and required manufacturing capabilities.	(Koufteros, Vonderembse, & Doll, 2002; Perrow, 1967; Salvador, 2007; Van de Ven & Delbecq, 1974)
Number of variables in decision making	Introduction of new products	Development of new products increases the need for information processing between units because it makes the manufacturing task less understandable; can new products be produced with the technology.	(Ettlie, Bridges, & O'Keefe, 1984; Fine, 1998; Takeuchi & Nonaka, 1986)
	Introduction of new technology	Development of new technology increases the need for information processing between units because it makes the manufacturing task less understandable; for example can products be produced with the new technology according to required timelines and quality.	(Allen, Tushman, & Lee, 1979; Ettlie et al., 1984; Olson, Walker, Ruekert, & Bonner, 2001)
Reciprocal task interdependence	Reciprocal task interdependence	Interdependence creates information processing needs because the organizational unit is dependent on the actions and outcomes of others. In particular reciprocal interdependence (consequently also team interdependence) creates integration needs.	(Galbraith, 1977; Thompson, 1967; Van de Ven et al., 1976)

sented in Table 4. The factor scores were then calculated with regression method (Bollen, 1989: 305; Hair, Anderson, Tatham, & Black, 1998: 119).

For multi-item scales, CFA was used to examine validity and reliability. Unidimensionality is established by a good fit of a single-factor model using CFA. Convergent validity is established by high and statistically significant factor loadings in the CFA model (average of loadings is 0.72 and all loadings are statistically significant at $p = 0.05$) (Bagozzi & Yi, 1991). Discriminant validity is established by low inter-construct correlations (CFA is conducted simultaneously for integration constructs and for task characteristics and the correlations are at maximum 0.27) (Bagozzi et al., 1991; O'Leary-Kelly & Vokurka, 1998). Finally, reliability is examined by calculating composite reliabilities (Hair et al., 1998). All composite reliabilities well exceed the threshold of 0.70 (see Table 4). Hence, we can conclude that unidimensionality, convergent and discriminant validities as well as reliability are adequate.

Identification of Internal Integration Strategies

In order to identify internal integration strategies, cluster analysis (Hair et al., 1998; Ketchen & Shook, 1996) was used. Using cluster analysis, the manufacturing organizations were classified based on their use of the five integration mechanisms. All variables were standardized (Ketchen & Shook, 1996). A two-step cluster procedure was conducted (for similar approach, see Bozarth et al., 2009; Zhao, Yeung, & Lee, 2004). First, we conducted hierarchical cluster analysis to determine the number of clusters. The hierarchical cluster analysis suggested four clusters. The decision was based on the changes in the agglomeration coefficient (Ketchen & Shook, 1996); the percentage in the change of the agglomeration coefficient

Table 4. Operationalization of multi-item scales

Construct, MSA, CR – Items, respondents		Loading
Centralization of decision making(MSA = 0.78, CR = 0.87) (based on Hage and Aiken (1967))		
	CE1 Even small matters have to be referred to someone higher up for a final answer.	.90
	CE2 This plant is a good place for a person who likes to make his own decisions. (reversed item)	.40
	CE3 Any decision I make has to have my boss's approval.	.85
	CE4 There can be little action taken here until a supervisor approves a decision.	.74
	Respondents: direct labor, HR manager, supervisor	
Cross-functional job rotation (MSA = 0.77, CR = 0.77)		
	JR1 Frequent rotation of managers between functions is normal practice in this plant.	.73
	JR2 Managers permanently specialize in one function at our plant. (reversed item)	.60
	JR3 Most of the managers here have had positions in more than one function.	.80
	JR4 Our managers have not worked outside of their own areas, for the most part. (reversed item)	.77
	Respondents: HR manager, plant manager, plant superintendent	
Plant-level incentives (MSA = 0.85, CR = 0.89)		
	IC1 Our incentive system encourages us to vigorously pursue plant objectives.	0.91
	IC2 The incentive system at this plant is fair at rewarding people who accomplish plant objectives.	0.90
	IC3 Our reward system really recognizes the people who contribute the most to our plant.	0.72
	IC4 The incentive system at this plant encourages us to reach plant goals.	0.93
	IC5 Our incentive system is at odds with our plant goals.	0.58
	Respondents: product development team member, process engineer, plant supervisor	
Product modularity (MSA = 0.77, CR = 0.96)		
	MO1 Our products are modularly designed, so they can be rapidly built by assembling modules.	.66
	MO2 We have defined product platforms as a basis for future product variety and options.	.54
	MO3 Our products are designed to use many common modules.	.89
	MO4 When we make two products that differ by only a specific feature, they generally require only one different subassembly/component.	.36
	MO5 We do not use common assemblies and components in many of our products. (reversed item)	.60
	Respondents: info systems manager, product development team member, process engineer	
Introduction of new manufacturing technology (MSA = 0.78, CR = 0.97)		
	NMT1 We pursue long-range programs, in order to acquire manufacturing capabilities in advance of our needs.	.62
	NMT2 We make an effort to anticipate the potential of new manufacturing practices and technologies.	.80
	NMT3 Our plant stays on the leading edge of new technology in our industry.	.63
	NMT4 We are constantly thinking of the next generation of manufacturing technology.	.85
	Respondents: process engineer, plant manager, plant superintendent	

was highest when the number of clusters changed from four to three, which suggests that a four-cluster solution yields the best result. Second, a non-hierarchical cluster analysis was conducted to generate the cluster results, setting the number of clusters to four. The cluster centroids, representing the mean values for each of the cluster variables, are presented in Table 6. The mean values represent the general characteristics of the clusters. We named the identified four clusters as centralizers (Group 1), information integrators

Table 5. Operationalization of single-item measures

Information systems	
	Are the following application areas supported by software: (i) demand planning, (ii) design (CAD, CAE), (iii) project management, and (iv) groupware tools? *Respondent:* info systems manager *Indicator:* A summated scale of the above items (0 = none of the applications is supported by software,…, 4 = all the four application areas are supported by software), standardized
Cross-unit teams	
	Assess the following statement (Scale: 1 = Strongly disagree,…, 7 = Strongly agree): We work in teams, with members from a variety of areas (marketing, manufacturing, etc.) to introduce new products. *Respondents:* product development team member, process engineer, supervisor
Customization of orders	
	Identify the importance of the goal rapid customization of orders (Scale: 1=Least important, …, 5=Absolutely crucial) *Respondents:* process engineer, plant manager, plant superintendent
Introduction of new products	
	What is the percentage of the plant sales that come from products introduced in the five years prior to the data collection? *Respondent:* process engineer
Reciprocal task interdependence	
	The product development process can be described by the following four stages: (1), Concept development/idea generation, (2) Product planning/technical and market feasibility, (3) Detailed design development and prototypes, (4) Manufacturing process development/pilot production Check the statement below that most closely describes your product development process: (i) The four stages were sequential, (ii) The four stages were sequential with some overlap, (iii) The four stages had significant overlap, (iv) The four stages were carried out simultaneously. *Respondent:* product development team member *Indicator:* Dichotomous (0 = sequential interdependence (categories 1 and 2), 1 = reciprocal interdependence (categories 3 and 4))

(Group 2), informal centralizers (Group 3) and intensive lateral integrators (Group 4).

After the cluster solution was generated, one-way analysis of variance (ANOVA) was conducted to compare the clusters. The results of ANOVA (Table 6) suggests that the use of all five integration mechanisms is different in the four integration strategies (p-value = 0.000). The Scheffe-test shows that 24 pairs, out of the 30 possible combinations of the five integration mechanism constructs, are highly significant (p

Table 6. Cluster centroids for the four-cluster solution

		Integration mechanisms				
Cluster		**Centralization**	**Information systems**	**Cross-unit job rotation**	**Cross-unit teams**	**Plant-level Incentives**
1	Centralizers	0.448	-0.654	-0.936	-0.529	-0.600
2	Information integrators	-0.347	0.660	-0.261	0.331	-0.028
3	Informal centralizers	0.276	-0.277	0.956	-0.248	0.335
4	Intensive lateral integrators	-1.453	0.393	0.768	1.199	1.039
F-statistic		33.272**	30.837**	97.420**	27.816**	26.499**

** p < 0.01; * p < 0.05; † p < 0.10

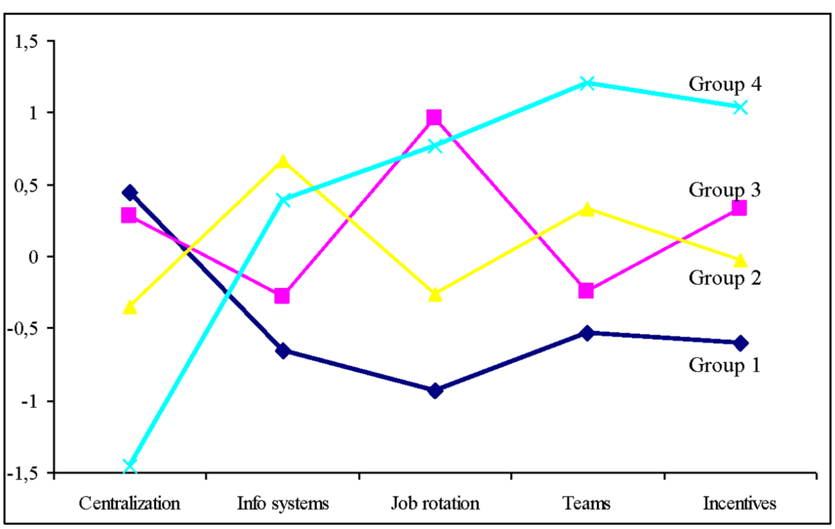
< 0.05) giving further support to the cluster solution.

The four clusters can be characterized as follows.

- Group 1: *Centralizers*. As Table 6 shows, the first group of plants is high on centralization of decision making and simultaneously low on the use of all other integration mechanisms. Hence, activities in these organizations are coordinated in traditional vertical ways by centralizing decision making to higher levels in the organization, and significantly less attention is put on managing coordination in a horizontal manner. Out of the 210 organizations, 62 (26%) organizations belong to this group.
- Group 2: *Information integrators*. The second group of organizations is low on centralization. Integration in this group is managed by heavy reliance on information systems and moderate reliance of cross-functional teams. Hence, rather than coordinating activities vertically, these organizations emphasize information processing across units by implementing information systems to support that and by establishing cross-unit teams. Out of the 210 plants, 55 (23%) belong to this group.
- Group 3: *Informal centralizers*. The third group of organizations is moderately centralized. However, instead of relying purely on centralization (cf. Group 1), this group of organizations has also implemented lateral mechanisms to manage the integration challenge. In particular, heavy emphasis has been put on cross-unit job rotation and moderate emphasis on incentives. This is the largest group, with 71 (30%) organizations out of the 210 belonging to this group.
- Group 4: *Intensive lateral integrators*. The last group of organizations is very low on centralization. In these organizations, activities are coordinated by relying heavily on all lateral mechanisms, including information systems, cross-unit job rotation, cross-unit teams, and incentives. Out of 210 organizations, 22 (9%) belong to this group.

The four-cluster solution is graphically presented in the following Figure 1.

Figure 1. The four-cluster solution



To conclude the cluster analysis, the results suggest that manufacturing organizations differ in terms of the emphasis they put on internal integration; rather than all organizations moving towards lateral integration, a significant amount of organizations also rely on more vertically focused integration strategies. The analysis points out four different integration strategies: centralizers, information integrators, informal centralizers, and intensive lateral integrators. Although the results are exploratory in nature, the findings are important and interesting because prior research has paid less attention to integration strategies. Furthermore, prior research looking at configurations has focused on the extent of integration (Das et al., 2006; Flynn et al., In press; Frohlich & Westbrook, 2001), rather than looking at how firms actually achieve integration with different combinations of managerial tools. This also makes it difficult to compare our findings with the configurations identified in the prior research. To a limited extent, it is possible to compare the configurations identified in this research with the overall organizational level structures proposed by Mintzberg (1980). Although he developed his configurations based on somewhat different variables, our centralizers closely resemble the simple structure proposed by Mintzberg (1980), being high on centralization and low on lateral relations and thus relying on coordination by direct supervision. Our intense lateral integrators, on the other hand, resemble Mintzberg's (1980) adhocracy, being low on centralization and high on reliance on lateral ways for coordinating activities.

Linking Integration Strategies to Task Characteristics

Building on contingency theory and the information processing perspective, the underlying assumption in this research is that the use of the identified integration strategies would differ depending on the characteristics of the organizational task. In the following, we assess whether the identified groups differ in terms of task characteristics. We used Analysis of Variance (ANOVA) to identify the potential differences in the characteristics of the task faced by the focal organization across the four clusters. The results of ANOVA are presented in Table 7.

The results suggest that there are some significant differences in the task characteristics across the clusters. However, the results are not very strong. In particular, significant differences are found in number of variables in cross-functional decision making, operationalized as the focus on introduction of new products and intro-

Table 7. Results of ANOVA assessing the differences in the task characteristics across the four clusters

	Group 1 Centralizers (N = 62)	Group 2 Information integrators (N = 55)	Group 3 Informal centralizers (N = 71)	Group 4 Intensive lateral integrators (N = 22)	F-statistic (Sig.)
Customization	-0.057	0.126	0.041	-0.285	0.990 (0.199)
Modularity	0.094	-0.127	-0.114	0.294	1.414 (0.120)
Introduction of new products	-0.166 (2)	0.345 (1, 3)	-0.110 (2)	0.021	2.375 (0.036)
Introduction of new tech.	-0.551 (2, 3, 4)	0.237 (1)	0.093 (1)	0.550 (1)	11.216 (0.000)
Reciprocal interdependence	0.509	0.320	0.448	0.571	1.827 (0.072)

Numbers in parentheses indicate the cluster(s) from which that particular cluster is significantly different at p<0.10

duction of new technology ($p < 0.05$) and reciprocal task interdependence ($p < 0.10$). Task variability (operationalized as customization of orders) and task difficulty (operationalized as product modularity) did not reveal differences across the clusters. The results suggest that centralizers are significantly lower than information integrators, in terms of introduction of new products and technology. Centralizers are also lower than informal centralizers and intensive lateral integrators, in terms of introduction of new technology. Information integrators, on the other hand, are higher than informal centralizers in terms of introduction of new products.

Building on the contingency theoretical perspective to integration (Galbraith, 1973; Lawrence & Lorsch, 1967b; Thompson, 1967), we suggested that, due to the various costs related to the implementation and use of integration mechanisms, we cannot assume that more emphasis on integration mechanisms would always be beneficial. In line with this, it is not suggested that there would be cumulative performance gains for organizations when moving from one configuration to another more complex one (cf. capability configurations presented by Narasimhan, Swink & Kim, 2005). Rather, we hypothesized that the effective integration could be related to the organization's task characteristics. Taking the classical contingency theory arguments to the configuration perspective, we could assume that there would be at least some difference in the use of the strategies so that, for example, centralizers have lower information processing requirements than intensive lateral integrators. The results suggest some differences in the use of different integration strategies, depending on the task characteristics. However, the results do not really support this, except for the case of introduction of new technology; the stronger the emphasis on introduction of new technology, the less centralizers are found and rather the organizations have implemented intensive lateral integration strategies. Significantly more research is still needed on addressing when different strategies are used and when they are effective.

Cross-Country Analysis of Integration Strategies

Additionally, our aim in this research was to explore how the use of the different integration strategies varies across countries. Understanding potential differences across countries would be especially managerially relevant and helpful in designing effective integration efforts, for example, in different units of a global manufacturing company. We used ANOVA to assess differences in the use of various integration strategies across countries. The results are presented in Table 8. The table includes both the number of plants, as well as the percentage of plants (in parenthesis) representing each strategy within a country.

The Pearson chi-square suggests that there are statistically significant differences in the use of integration strategies across the countries (p-value = 0.000). For example, US plants tend to be either information integrators or informal centralizers. Most of German, Finnish, and Swedish tend to have a somewhat different profile and be either centralizers or informal centralizers. Most of Korean plants are centralizers or information integrators, and Japanese plants tend to be information integrators. When looking at the use of different integration strategies across countries, the results suggest that Finnish, Austrian and German plants follow relatively more often the intensive lateral integration strategy, whereas none of the Japanese or Korean plants in the sample followed that integration strategy.

One explanation for the different use of integration strategies can be found in the differences between the national cultural environments in different countries. In the organization design literature, it has been reported that, due to, for example, the cultural dimensions of individualism–collectivism (Hofstede, 1980), centralization is higher in Asian countries than in Scandinavia.

Table 8. Differences of integration strategies across countries

		Group 1 Centralizers (N = 62)	Group 2 Information integrators (N = 55)	Group 3 Informal centralizers (N = 71)	Group 4 Intensive lateral integrators (N = 22)	Total
Country	Austria	2 (11%)	4 (21%)	9 (47%)	4 (21%)	19
	Finland	8 (27%)	6 (20%)	10 (33%)	6 (20%)	30
	Germany	12 (30%)	2 (5%)	19 (48%)	7 (17%)	40
	Italy	13 (50%)	4 (15%)	9 (35%)	0 (0%)	26
	Japan	5 (18%)	18 (67%)	4 (15%)	0 (0%)	27
	South Korea	9 (39%)	12 (52%)	2 (9%)	0 (0%)	23
	Sweden	9 (41%)	1 (4%)	9 (41%)	3 (14%)	22
	United States	4 (17%)	8 (35%)	9 (39%)	2 (9%)	23
	Total	55	62	71	22	210

This could give an explanation for why Korean and Japanese plants follow a centralization strategy rather than being intensive lateral integrators. More research is clearly needed in this area, both in the context of external and internal integration.

CONCLUSION

This chapter has discussed supply chain integration by focusing on internal integration within a manufacturing unit. Based on an international survey of manufacturing plants, our aim was to identify internal integration strategies. Our empirical analysis revealed the existence of four separate groups of plants, which differ in terms of managerial tools that they have implemented to achieve integration within their organization. These were named centralizers, information centralizers, informal integrators, and intensive lateral integrators. We further suggested, based on contingency theory and the information processing perspective, that none of the integration strategies is superior in overall terms. Rather, effective use of integration strategy depends on the organizational context. This argument builds on the idea that implementation and use of integration mechanisms poses benefits, but also costs, to an organization. We analyzed the use of different integration strategies under varying task characteristics, as well as across countries.

This chapter is highly exploratory in nature. The research presented in this chapter extends prior research on supply chain integration in several different ways. First, this chapter complements prior research on supply chain integration by focusing on internal integration within a manufacturing organization. The majority of the prior work on supply chain integration has focused on discussing external relationships between the focal firm and its suppliers and customers. Second, the chapter also complements prior research by discussing how companies actually achieve integrated behavior in different contexts. The majority of the prior work has focused on looking at the effects of integration on performance, giving less emphasis on how integration is managed. The chapter has, in particular, contributed to the research by assessing different integration strategies of manufacturing firms. Finally, and most importantly, this chapter has emphasized the potential costs of integration efforts. The majority of the prior research builds on the assumption that supply chain integration is highly and equally important to all organizations.

We question this assumption and suggest that integration efforts pose significant and varying costs to the focal organization. Hence, in order to operate effectively and efficiently, organizations need to assess whether integration is needed and how it can be managed in an efficient way.

In addition to the academic implications, the research is also managerially relevant. We identified internal integration strategies can work as a framework for managers when analyzing integration within their organization. The chapter also draws managers' attention to the costs of integration. The analysis of contextual, as well as national, differences in the use of the internal integration strategies can also help managers when designing their manufacturing units in the global economy, for example, which integration strategies could be more applicable in different circumstances and environments.

Limitations

Although contributing to the discussion on supply chain integration, this research is not without limitations. First, this research used five integration mechanisms, in order to identify possible integration strategies. Although these mechanisms represent different categories of integration mechanisms, and subsequently different capacity to facilitate integration and different costs, there are number of other mechanisms which could be included in subsequent analysis. Second, some operationalizations, especially single-item constructs, are problematic. Third, contingency theory has sometimes been criticized for its cross-sectional nature (e.g., Dewar & Hage, 1978), and longitudinal studies have been called for (Aldrich, 2001). Porter (1991), however, argues that cross-sectional studies are logically prior to longitudinal ones, justifying the cross-sectional research approach. Hence, it is important to first build theoretical arguments and conduct empirical analysis with cross-sectional data and only later develop and test those arguments in a dynamic context. Finally, common method bias is a potential limitation (for a detailed discussion, see Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Potential sources of common method bias include for example social desirability, leniency bias, and item ambiguity.

Future Research

Empirical analysis revealed four different integration strategies for internal integration within manufacturing organizations. The analysis was highly exploratory in nature. Although the clusters are interesting and theoretically rational from a contingency perspective, they are still highly preliminary in nature. Also analysis of task characteristics under which different strategies are used yielded only weak results. Hence, future research could engage in more detailed empirical, but especially theoretical, analysis of configurations of integration mechanisms. What kind of strategies could be expected to be found? In which situations are certain strategies more effective and used than others? Which factors determines the viable integration strategy?

This chapter focuses on internal supply chain integration in manufacturing companies. The ideas and discussion presented in the chapter raise ideas for research regarding external integration in the supply network. Future research could address how decision making is managed across firms which have different goals and interests; what different kinds of integration strategies are used to manage the external integration challenge? Are similar types of strategies found in that context? Is there a link between the internal integration strategy and external strategy used?

This research has focused on analyzing integration strategies. In addition, this research has also explored whether the use of strategies differs across different conditions. Hence, the focus has been on understanding the internal design of organizations giving less emphasis to the performance effects of organization design. Future research could address how internal integration strategies are related to performance; how are integration strategies related to performance in different contexts? Which integration strategy provides the most benefits in terms of performance in various contexts? How do internal integration strategy and external integration support each other in improving organizational performance?

REFERENCES

Adler, P. S. (1995). Interdepartmental Interdependence and Coordination: The Case of the Design/Manufacturing Interface. *Organization Science*, *6*(2), 147–167. doi:10.1287/orsc.6.2.147

Aldrich, H. E. (2001). Who Wants to Be an Evolutionary Theorist? Remarks on the Occasion of the Year 2000 OMT Distinguished Scholarly Career Award Presentation. *Journal of Management Inquiry*, *10*(2), 115–127. doi:10.1177/1056492601102004

Allen, T. J., Tushman, M. L., & Lee, D. M. S. (1979). Technology Transfer as a Function of Position in the Spectrum from Research through Development to Technical Services. *Academy of Management Journal*, *22*(4), 694–708. doi:10.2307/255809

Bagchi, P. K., & Skjoett-Larsen, T. (2002). Integration of Information Technology and Organizations in Supply Chain. *The International Journal of Logistics Management*, *14*(1), 89–108. doi:10.1108/09574090310806477

Bagozzi, R. P., & Yi, Y. (1991). Multitrait-Multimethod Matrices in Consumer Research. *The Journal of Consumer Research*, *17*(4), 426–439. doi:10.1086/208568

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, *36*(3), 421–458. doi:10.2307/2393203

Bajaj, A., Kekre, S., & Srinivasan, K. (2004). Managing NPD: Cost and Schedule Performance in Design and Manufacturing. *Management Science*, *50*(4), 527–536. doi:10.1287/mnsc.1030.0177

Ballou, R. H., Gilbert, S. M., & Mukherjee, A. (2000). New Managerial Challenges from Supply Chain Opportunities. *Industrial Marketing Management*, *29*(1), 7–18. doi:10.1016/S0019-8501(99)00107-8

Barki, H., & Pinsonneault, A. (2005). A Model of Organizational Integration, Implementation Effort, and Performance. *Organization Science*, *16*(2), 165–179. doi:10.1287/orsc.1050.0118

Barnard, C. I. (1938). *The Functions of the Executive*. Cambridge, MA: Harvard University Press.

Bartlett, C. A., & Ghoshal, S. (1989). *Managing Across Borders - The Transnational Solution*. Harvard Business School Press.

Bensaou, M., & Venkatraman, N. (1995). Configurations of Interorganizational Relationships: A Comparison between U.S. and Japanese Automakers. *Management Science*, *41*(9), 1471–1492. doi:10.1287/mnsc.41.9.1471

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Bozarth, C. C., Warsing, D. P., Flynn, B. B., & Flynn, E. J. (2009). The Impact of Supply Chain Complexity on Manufacturing Plant Performance. *Journal of Operations Management*, *27*(1), 78–93. doi:10.1016/j.jom.2008.07.003

Burns, T., & Stalker, G. M. (1961). *The Management of Innovation* (3rd ed.). London, UK: Tavistock Publications.

Chen, I. J., & Paulraj, A. (2004). Understanding Supply Chain Management: Critical Research and a Theoretical Framework. *International Journal of Production Research*, *42*(1), 131–163. doi:10.1080/00207540310001602865

Child, J. (1972). Organization Structure and Strategies of Control: A Replication of the Aston Study. *Administrative Science Quarterly*, *17*(2), 163–177. doi:10.2307/2393951

Child, J. (1973a). Predicting and Understanding Organization Structure. *Administrative Science Quarterly*, *18*(2), 168–185. doi:10.2307/2392061

Child, J. (1973b). Strategies for Control and Organizational Behavior. *Administrative Science Quarterly*, *18*(1), 1–17. doi:10.2307/2391923

Child, J. (1975). Managerial and Organizational Factors Associated with Company Performance - Part II. A Contingency Analysis. *Journal of Management Studies*, *12*(1), 12–27. doi:10.1111/j.1467-6486.1975.tb00884.x

Clark, K. B. (1989). Project Scope and Project Performance: The Effect of Parts Strategy and Supplier Involvement on Product Development. *Management Science*, *35*(10), 1247–1263. doi:10.1287/mnsc.35.10.1247

Cousins, P. D., & Menguc, B. (2006). The Implications of Socialization and Integration in Supply Chain Management. *Journal of Operations Management*, *24*(5), 604–620. doi:10.1016/j.jom.2005.09.001

Daft, R. L., & Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, *32*(5), 554–571. doi:10.1287/mnsc.32.5.554

Das, A. J., Narasimhan, R., & Talluri, S. (2006). Supplier Integration - Finding an Optimal Configuration. *Journal of Operations Management*, *24*(5), 563–582. doi:10.1016/j.jom.2005.09.003

Dewar, R., & Hage, J. (1978). Size, Technology, Complexity, and Structural Differentiation: Toward a Theoretical Synthesis. *Administrative Science Quarterly*, *23*(1), 111–136. doi:10.2307/2392436

Dill, W. R. (1958). Environment as an Influence on Managerial Autonomy. *Administrative Science Quarterly*, *2*(4), 409–443. doi:10.2307/2390794

Drazin, R., & Van de Ven, A. H. (1985). Alternative Forms of Fit in Contingency Theory. *Administrative Science Quarterly*, *30*(4), 514–539. doi:10.2307/2392695

Edström, A., & Galbraith, J. R. (1977). Transfer of Managers as a Coordination and Control Strategy in Multinational Organizations. *Administrative Science Quarterly*, *22*(2), 248–263. doi:10.2307/2391959

Ellinger, A. E., Daugherty, P. J., & Keller, S. B. (2000). The Relationship between Marketing/Logistics Interdepartmental Integration and Performance in U.S. Manufacturing Firms: An Empirical Study. *Journal of Business Logistics*, *21*(1), 1–22.

Ettlie, J. E. (1995). Product-Process Development Integration in Manufacturing. *Management Science*, *41*(7), 1224–1237. doi:10.1287/mnsc.41.7.1224

Ettlie, J. E., Bridges, W. P., & O'Keefe, R. D. (1984). Organization Strategy and Structural Differences for Radical versus Incremental Innovation. *Management Science*, *30*(6), 682–695. doi:10.1287/mnsc.30.6.682

Ettlie, J. E., & Reza, E. M. (1992). Organizational Integration and Process Innovation. *Academy of Management Journal*, *35*(4), 795–827. doi:10.2307/256316

Fawcett, S. E., & Magnan, G. M. (2002). The Rhetoric and Reality of Supply Chain Management. *International Journal of Physical Distribution & Logistics Management*, *32*(5), 339–361. doi:10.1108/09600030210436222

Fine, C. H. (1998). *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*. Reading, MA: Perseus Books.

Fine, C. H. (2000). Clockspeed-Based Strategies for Supply Chain Design. *Production and Operations Management*, *9*(3), 213–221. doi:10.1111/j.1937-5956.2000.tb00134.x

Flynn, B. B., Huo, B., & Zhao, X. (in press). The Impact of Supply Chain Integration on Performance: A Contingency and Configuration Approach. *Journal of Operations Management*.

Forza, C. (2002). Survey Research in Operations Management: A Process-Based Perspective. *International Journal of Operations & Production Management*, *22*(2), 152–194. doi:10.1108/01443570210414310

Forza, C., & Salvador, F. (2002). Managing for Variety in the Order Acquisition and Fulfillment Process: The Contribution of Product Configuration Systems. *International Journal of Production Economics*, *76*(1), 87–98. doi:10.1016/S0925-5273(01)00157-8

Frohlich, M. T., & Westbrook, R. (2001). Arcs of Integration: An International Study of Supply Chain Strategies. *Journal of Operations Management*, *19*(2), 185–200. doi:10.1016/S0272-6963(00)00055-3

Galbraith, J. R. (1973). *Designing Complex Organizations*. Reading, MA: Addison-Wesley.

Galbraith, J. R. (1977). *Organization Design*. Reading, MA: Addison-Wesley.

Galbraith, J. R. (1994). *Competing with Flexible Lateral Organizations* (2nd ed.). Reading, MA: Addison-Wesley.

Galbraith, J. R. (2000). *Designing the Global Corporation*. San Francisco: Jossey-Bass.

Galbraith, J. R. (2002). *Designing Organizations - An Executive Guide to Strategy, Structure, and Process*. (New and revised ed.). San Francisco: Jossey-Bass.

Germain, R., Claycomb, C., & Dröge, C. (2008). Supply Chain Variability, Organizational Structure, and Performance: The Moderating Effect of Demand Unpredictability. *Journal of Operations Management*, *26*(5), 557–570. doi:10.1016/j.jom.2007.10.002

Germain, R., & Iyer, K. N. S. (2006). The Interaction of Internal and Downstream Integration and its Association with Performance. *Journal of Business Logistics*, *27*(2), 29.

Gerwin, D. (1981). Relationships between Structure and Technology. In: P. C. Nystom & W. H. Starbuck (Eds.), *Handbook of Organizational Design*, Vol. 2, pp. (3-38). New York: Oxford University Press.

Gimenez, C., & Ventura, E. (2005). Logistics-Production, Logistics-Marketing and External Integration - Their Impact on Performance. *International Journal of Operations & Production Management*, *25*(1), 20–38. doi:10.1108/01443570510572222

Grandori, A., & Soda, G. (1995). Inter-firm Networks: Antecedents, Mechanisms, and Forms. *Organization Studies*, *16*(2), 183–214. doi:10.1177/017084069501600201

Hage, J., & Aiken, M. (1967). Relationship of Centralization to Other Structural Properties. *Administrative Science Quarterly*, *12*(1), 72–92. doi:10.2307/2391213

Hage, J., Aiken, M., & Marrett, C. B. (1971). Organization Structure and Communications. *American Sociological Review*, *36*(5), 860–871. doi:10.2307/2093672

Hair, J. F. Jr, Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hofstede, G. (1980). *Culture's Consequences - International Differences in Work-Related Values*. Beverly Hills, CA: Sage Publications.

Kahn, K. B., & Mentzer, J. T. (1996). Logistics and Interdepartmental Integration. *International Journal of Physical Distribution & Logistics Management*, *26*(8), 6–14. doi:10.1108/09600039610182753

Ketchen, D. J. Jr, & Shook, C. L. (1996). The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strategic Management Journal*, *17*(6), 441–458. doi:10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

Ketchen, D. J. Jr, Thomas, J. B., & Snow, C. C. (1993). Organizational Configurations and Performance: A Comparison of Theoretical Approaches. *Academy of Management Journal*, *36*(6), 1278–1313. doi:10.2307/256812

Koufteros, X. A., Cheng, T. C. E., & Lai, K.-H. (2007). "Black-box" and "Grey-box" Supplier Integration in Product Development: Antecedents, Consequences and the Moderating Role of Firm Size. *Journal of Operations Management*, *25*(4), 847–870. doi:10.1016/j.jom.2006.10.009

Koufteros, X. A., Vonderembse, M. A., & Doll, W. J. (2002). Integrated Product Development Practices and Competitive Capabilities: The Effects of Uncertainty, Equivocality, and Platform Strategy. *Journal of Operations Management*, *20*(4), 331–355. doi:10.1016/S0272-6963(02)00018-9

Koufteros, X. A., Vonderembse, M. A., & Jayaram, J. (2005). Internal and External Integration for Product Development: The Contingency Effect of Uncertainty, Equivocality, and Platform Strategy. *Decision Sciences*, *36*(1), 97–133. doi:10.1111/j.1540-5915.2005.00067.x

Lambert, D. M., Robeson, J. F., & Stock, J. R. (1978). An Appraisal of the Integrated Physical Distribution Management Concept. *International Journal of Physical Distribution and Materials Management*, *9*(1), 74–88.

Lawrence, P. R. (1981). Organization and Environment Perspectives. In: A. H. Van de Ven & W. F. Joyce (Eds.), *Perspectives on Organization Design and Behavior*, pp. (311-337). New York: Wiley.

Lawrence, P. R., & Lorsch, J. W. (1967a). Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*, *12*(1), 1–47. doi:10.2307/2391211

Lawrence, P. R., & Lorsch, J. W. (1967b). *Organization and Environment - Managing Differentiation and Integration*. Boston: Harvard Business School Press.

March, J. G., & Simon, H. A. (1958). *Organizations* (2nd ed.). Cambridge, MA: Blackwell.

Martinez, J. I., & Jarillo, J. C. (1989). The Evolution of Research on Coordination Mechanisms in Multinational Corporations. *Journal of International Business Studies*, *20*(3), 489–514. doi:10.1057/palgrave.jibs.8490370

Meyer, A. D., Tsui, A. S., & Hinings, C. R. (1993). Configurational Approaches to Organizational Analysis. *Academy of Management Journal*, *36*(6), 1175–1195. doi:10.2307/256809

Miles, R. E., & Snow, C. C. (2007). Organization Theory and Supply Chain Management: An Evolving Research Perspective. *Journal of Operations Management*, *25*(2), 459–463. doi:10.1016/j.jom.2006.05.002

Mintzberg, H. (1980). Structure in 5's: A Synthesis of the Research on Organizational Design. *Management Science*, *26*(3), 322–341. doi:10.1287/mnsc.26.3.322

Narasimhan, R., & Kim, S. W. (2002). Effect of Supply Chain Integration on the Relationship Between Diversification and Performance: Evidence from Japanese and Korean Firms. *Journal of Operations Management*, *20*(3), 303–323. doi:10.1016/S0272-6963(02)00008-6

Narasimhan, R., Swink, M., & Kim, S. W. (2005). An Exploratory Study of Manufacturing Practice and Performance Interrelationships - Implications for Capability Progression. *International Journal of Operations & Production Management*, *25*(9-10), 1013–1033. doi:10.1108/01443570510619509

O'Leary-Kelly, S. W., & Flores, B. E. (2002). The Integration of Manufacturing and Marketing/sales Decisions: Impact on Organizational Performance. *Journal of Operations Management*, *20*(3), 221–240. doi:10.1016/S0272-6963(02)00005-0

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The Empirical Assessment of Construct Validity. *Journal of Operations Management*, *16*(4), 387–405. doi:10.1016/S0272-6963(98)00020-5

Olson, E. M., Walker, O. C. Jr, Ruekert, R. W., & Bonner, J. M. (2001). Patterns of Cooperation during New Product Development among Marketing, Operations and R&D: Implications for Project Performance. *Journal of Product Innovation Management*, *18*(4), 258–271. doi:10.1016/S0737-6782(01)00091-1

Pagell, M. (2004). Understanding the Factors that Enable and Inhibit the Integration of Operations, Purchasing, and Logistics. *Journal of Operations Management*, *22*(5), 459–487. doi:10.1016/j.jom.2004.05.008

Perrow, C. (1967). A Framework for the Comparative Analysis of Organizations. *American Sociological Review*, *32*(2), 194–208. doi:10.2307/2091811

Pierce, J. L., & Delbecq, A. L. (1977). Organization Structure, Individual Attitudes and Innovation. *Academy of Management Review*, *2*(1), 27–37. doi:10.2307/257602

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. doi:10.1037/0021-9010.88.5.879

Porter, M. E. (1985). *Competitive Advantage - Creating and Sustaining Superior Performance*. New York: The Free Press.

Porter, M.E. (1991). Towards a Dynamic Theory of Strategy. *Strategic Management Journal* 12 (Special Issue: Fundamental Reserch Issues in Strategy and Economics, Winter), 95-117.

Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. (1968). Dimensions of Organization Structure. *Administrative Science Quarterly*, *13*(1), 65–105. doi:10.2307/2391262

Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. (1969). The Context of Organization Structures. *Administrative Science Quarterly*, *14*(1), 91–114. doi:10.2307/2391366

Rivkin, J. W. (2001). Reproducing Knowledge: Replication Without Imitation at Moderate Complexity. *Organization Science*, *12*(3), 274–293. doi:10.1287/orsc.12.3.274.10106

Rodriques, A. M., Stank, T. P., & Lynch, D. F. (2004). Linking Strategy, Structure, Process, and Performance in Integrated Logistics. *Journal of Business Logistics*, *25*(2), 65.

Rosenzweig, E. D., Roth, A. V., & Dean, J. W. Jr. (2003). The Influence of an Integration Strategy on Competitive Capabilities and Business Performance: An Exploratory Study of Consumer Products Manufacturers. *Journal of Operations Management*, *21*(4), 437–456. doi:10.1016/S0272-6963(03)00037-8

Salvador, F. (2007). Toward a Product System Modularity Construct: Literature Review and Reconceptualization. *IEEE Transactions on Engineering Management*, *54*(2), 219–240. doi:10.1109/TEM.2007.893996

Schroeder, R. G., & Flynn, B. B. (2001). *High Performance Manufacturing - Global Perspectives*. New York: John Wiley & Sons.

Snow, C. C., Miles, R. E., & Miles, G. (2006). The Configurational Approach to Organizational Design: Four Recommended Initiatives. In Burton, R. M., Eriksen, B., Håkonsson, D. D., & Snow, C. C. (Eds.), *Organization Design: The Evolving State-of-the-Art, pp* (pp. 3–18). New York: Springer Publishing.

Spina, G., Verganti, R., & Zotteri, G. (2002). Factors Influencing Co-design Adoption: Drivers and Internal Consistency. *International Journal of Operations & Production Management*, *22*(12), 1354–1366. doi:10.1108/01443570210452048

Stevens, G. C. (1989). Integrating the Supply Chain. *International Journal of Physical Distribution and Materials Management*, *19*(8), 3–8.

Swink, M. (2000). Technological Innovativeness as a Moderator of New Product Design Integration and Top Management Support. *Journal of Product Innovation Management*, *17*(3), 208–220. doi:10.1016/S0737-6782(00)00040-0

Swink, M., Narasimhan, R., & Wang, C. (2007). Managing Beyond the Factory Walls: Effects of Four Types of Strategic Integration on Manufacturing Plant Performance. *Journal of Operations Management*, *25*(1), 148–164. doi:10.1016/j.jom.2006.02.006

Takeuchi, H., & Nonaka, I. (1986). The New Product Development Game. *Harvard Business Review*, *64*(1), 137–146.

Thompson, J. D. (1967). *Organizations in Action - Social Science Bases of Administrative Theory*. (Transaction ed.). New Brunswick, NJ: Transaction Publishers.

Tushman, M. L. (1977). Special Boundary Roles in the Innovation Process. *Administrative Science Quarterly*, *22*(4), 587–605. doi:10.2307/2392402

Tushman, M. L., & Nadler, D. A. (1978). Information Processing as an Integrating Concept in Organizational Design. *Academy of Management Review*, *3*(3), 613–624. doi:10.2307/257550

Vachon, S., & Klassen, R. D. (2008). Environmental Management and Manufacturing Performance: The Role of Collaboration in the Supply Chain. *International Journal of Production Economics*, *111*(2), 299–315. doi:10.1016/j.ijpe.2006.11.030

Van de Ven, A. H., & Delbecq, A. L. (1974). A Task Contingent Model of Work-Unit Design. *Administrative Science Quarterly*, *19*(2), 183–197. doi:10.2307/2393888

Van de Ven, A. H., Delbecq, A. L., & Koenig, R. Jr. (1976). Determinants of Coordination Modes within Organizations. *American Sociological Review*, *41*(2), 322–338. doi:10.2307/2094477

Van der Vaart, T., & Van Donk, D. P. (2007). A Critical Review of Survey-Based Research in Supply Chain Integration. *International Journal of Production Economics*, *111*(1), 42–55. doi:10.1016/j.ijpe.2006.10.011

Vickery, S., Dröge, C., & Germain, R. (1999). The Relationship Between Product Customization and Organization Structure. *Journal of Operations Management*, *17*(4), 377–391. doi:10.1016/S0272-6963(98)00053-9

Vickery, S., Jayaram, J., Droge, C., & Calantone, R. (2003). The Effects of an Integrative Supply Chain Strategy on Customer Service and Financial Performance: An Analysis of Direct versus Indirect Relationship. *Journal of Operations Management*, *21*(5), 523–539. doi:10.1016/j.jom.2003.02.002

Wheelwright, S. C., & Clark, K. B. (1992). *Revolutionizing Product Development - Quantum Leaps in Speed, Efficiency, and Quality*. New York: Free Press.

Woodward, J. (1965). *Industrial Organization: Theory and Practice* (2nd ed.). London: Oxford University Press.

Zhao, X. D., Yeung, A. C. L., & Lee, T. S. (2004). Quality Management and Organizational Context in Selected Service Industries of China. *Journal of Operations Management*, *22*(6), 575–587. doi:10.1016/j.jom.2004.08.003

ADDITIONAL READING

Daft, R. L. (2004). *Organization Theory and Design* (8th ed.). Mason, OH: Thomson South-Western.

Donaldson, L. (1996). The Normal Science of Structural Contingency Theory. InS. R. Clegg & C. Hardy & W. R. Nord (Eds.), *Handbook of Organizational Studies*, pp. (57-76). London, UK: Sage Publications.

Galbraith, J. R. (1970). Environmental and Technological Determinants of Organizational Design. In Lorsch, J. W., & Lawrence, P. R. (Eds.), *Studies in Organization Design* (pp. 113–139). Homewood, IL: Richard D. Irwin and The Dorsey Press.

Ghemawat, P. (2001). Distance Still Matters - The Hard Reality of Global Expansion. *Harvard Business Review*, *79*(8), 137–147.

Ghoshal, S., & Gratton, L. (2002). Integrating the Enterprise. *Sloan Management Review*, *44*(1), 31–39.

Lee, H. L. (2002). Aligning Supply Chain Strategies with Product Uncertainties. *California Management Review*, *44*(3), 105–119.

McCann, J., & Galbraith, J. R. (1981). Interdepartmental Relations. In Nystom, P. C., & Starbuck, W. H. (Eds.), *Handbook of Organizational Design* (pp. 60–84). New York: Oxford University Press.

McCann, J. E., & Ferry, D. L. (1979). An Approach for Assessing and Managing Inter-Unit Interdependence. *Academy of Management Review*, *4*(1), 113–120. doi:10.2307/257411

Nadler, D. A., & Tushman, M. L. (1997). *Competing by Design - The Power of Organizational Architecture*. New York: Oxford University Press.

Stevens, G. C. (1989). Integrating the Supply Chain. *International Journal of Physical Distribution and Materials Management*, *19*(8), 3–8.

Trautmann, G., Turkulainen, V., & Hartmann, E. (2009). Integration in the Global Sourcing Organization - An Information Processing Perspective. *Journal of Supply Chain Management*, *45*(2), 57–74. doi:10.1111/j.1745-493X.2009.03163.x

Trent, R. J. (2004). The Use of Organizational Design Features in Purchasing and Supply Chain Management. *The Journal of Supply Chain Management*, *40*(3), 4–18. doi:10.1111/j.1745-493X.2004.tb00170.x

Turkulainen, V. (2008). *Managing Cross-functional Interdependencies - The Contingent Value of Integration*. Espoo, Finland: Helsinki University of Technology.

Van de Ven, A. H., & Drazin, R. (1985). The Concept of Fit in Contingency Theory. In Staw, B. M., & Cummings, L. L. (Eds.), *Research in Organizational Behavior, 7* (pp. 333–365). Greenwich, CT: JAI Press.

Van de Ven, A. H., & Walker, G. (1984). The Dynamics of Interorganizational Coordination. *Administrative Science Quarterly*, *29*(4), 598–621. doi:10.2307/2392941

KEY TERMS AND DEFINITIONS

Configuration: Any multidimensional constellation of conceptually distinct characteristics that commonly occur together.

External Integration: Integration in the context of management of relationships between formally separated organizations. These organizations can be for example in supplier-customer relationship or in other kinds of partnership.

Integration (Coordination): Both the state of unity in effort in the actions of various subunits, as well as the process of achieving that unity. In this research, the terms coordination and integration are used interchangeably.

Integration Mechanism: Any managerial tool for achieving integration and coordinated activities within an organization, including for example centralization, information systems, and lateral structures.

Integration Strategy: A combination (configuration, pattern) of individual integration mechanisms implemented and used by an organization in order to manage inter-organizational or intra-organizational relationships.

Internal Integration: Integration in the context of management of relationships between organizational units within an organization. These organizational units can be for example functionally related (different functional units), within the same function (for example manufacturing units located in different geographical locations), or in a vertical relationship (for example corporate office and manufacturing unit).

Supply Chain Integration: The degree to which a manufacturing firm collaborates with its supply chain partners and manages intra- and inter-organizational processes in a collaborative way. Managing inter-organizational relationships refers to external integration and intra-organizational relationships to internal integration.

Uncertainty: The difference between the amounts of information required to perform a task and the amount of information already possessed by the organization.

This work was previously published in Managing Global Supply Chain Relationships: Operations, Strategies and Practices, edited by Barbara Flynn, Michiya Morita and Jose Machuca, pp. 155-177, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 26
Equipment Replacement Decisions Models with the Context of Flexible Manufacturing Cells

Ioan Constantin Dima
Valahia University of Târgovişte, Romania

Janusz Grabara
Częstochowa University of Technology, Poland

Mária Nowicka-Skowron
Częstochowa University of Technology, Poland

ABSTRACT

The main problem of establishing equipment replacement decisions rules under specific conditions is to find decision variables that minimize total incurred costs over a planning horizon. Basically, the rules differ depending on what type of production type is used. For a batch production organization the suitable criterion is built on the principle of economies of scale. Proposed econometric models in this chapter are focused on a multiple machine replacement problem in flexible manufacturing cells with several machines for parts' processing, and industrial robots for manipulation and transportation of manufactured objects. Firstly, models for a simple case multiple machine replacement problems are presented. Subsequently, the more complicated case is considered where technological improvement is taken into account.

INTRODUCTION

Historically, development of production processes has passed from production structures in automatic rigid flow lines, efficient for mass and wide-range production, to flexible structures, especially efficient in low and medium-range production. Because manufacturing firms has to be flexible towards new market requirements, flexible production forms are increasingly seen as one of the most important manufacturing concepts. Currently, the trend in flexible manufacturing systems is toward small flexible manufacturing structures,

DOI: 10.4018/978-1-4666-1945-6.ch026

called flexible manufacturing cells (FMC). In this sense, two or more CNC machines are considered a flexible cell and two or more cells are considered a flexible manufacturing system (Groover, 2001). Flexible manufacturing cell, in general, allows the processing of pieces which are different in terms of shape and dimensions, in a determined range. This creates prerequisites for the accomplishment of variable products, under high yield conditions. Considerable savings are made because the utilizations increases, the processing time is shortened, the handling distances are reduced, intermediate storage expenses decrease, the area required for production is reduced, the process may be systematized, proper conditions for continuous work are created and direct expenses are reduced. However, the real occurrence of failures during the exploitation stage can markedly modify the FMC performances (Corbaa et al, 1997). For this reason, downtime of FMC has to be analyzed and its influence on the processing cost has to be pondered over. In addition, to ensure that manufacturing process is held to be competitive, upgrading or replacing of equipment due to rapid innovations in technology also has to be considered. In this context often encountered issues in production planning are: Should this equipment be replaced? If not now, then when? Usually, written equipment replacement policy, in which units are scheduled for replacement based on age and expected condition, contains answers on such questions. In this chapter we wish to show several econometric models that could inspire managers to develop their own specific tools in building of an effective equipment replacement policy.

THE PROBLEM STATEMENT

Theoretically, any equipment replacement decision would be made based on thorough modeling equipment deterioration and projected remaining life. Practical approaches to equipment replacement decisions are mostly based on subjective appraisal. But it is generally accepted that tools for equipment replacement decision create important element of repair/replacement policy. Such a policy provides guidance to production and economic manager regarding when to replace existing equipment or its part; how to conduct the acquisition process; and what should be done with the equipment being replaced. Then, the main importance of developing of equipment replacement decision models in production planning consist in establishing rules for the replacement of old equipment or its part(s) by new. The main problem of establishing the rules is to find decision variables that minimize total incurred costs over a planning horizon (Dehayem Nodem et al 2009). Basically, the rules differ depending on what type of production type is used. For batch production organizations suitable criteria are built on the principle of economies of scale, where the large fixed costs of production are depreciation-intensive because of huge capital investments made in high-volume operations and are spread over large production batch sizes in an effort to minimize the total unit costs of owning and operating the manufacturing system (Sullivan, 2002). When solving equipment/parts replacement problem within a flexible manufacturing cell, it is necessary to consider the impacts of the replacement decisions on all of the components of the system. Therefore, possible equipment stoppages due to wrong decision results at least in diminishing capacity or stopping the operations in a manufacturing cell. Accordingly, proposed methods are dedicated for a multiple machine replacement problem that is also characterized as a flexible flow shop problem. A parallel flow shop production concept is consisting of a number of production lines. Jobs in such work shop may be composed of a series of works, each requiring several machines (Jianhua and Fujimoto, 2003).

Firstly we will model a simple case multiple machine replacement problem that is characterized for a parallel flexible flow shop environment, in which no technological improvement in equipment is in concern. Then we will consider the more complicated case where we also take account of technological improvement.

RELATED WORK

Equipment replacement, as a specific field of knowledge and practice, has been extensively studied in the professional literature from the third decade of the 20th century (Castro et al 2009). Operations research approaches utilized in this domain are classified based on methods used to solve replacement problems, such as: integer programming (Hritonenko and Yatsenko, 2007), dynamic programming (Flynn and Chung, 2004), simulation techniques (Freeman, 1996) and Markov decision problems (Love et al, 2000).

Equipment replacement decision approaches related to this work can be divided into two basic types: parallel and series. The difference between these two categories is that in parallel models, the capacity of the system is simply the sum of the capacities of the individual assets and in the series – flow shop models, the minimum capacity assets in the series defines the capacity of the system (Hartman and Ban, 2002). The literature on parallel models is relatively rich. Among the many papers published on this topic, the interested reader may refer to Bean et al. (1994). As regards to series models, there is limited work. For instance, Tanchoco and Leung (1987), Suresh (1991, 1992) and Stinson and Khumawala (1987) present various approaches in which machines operate in series.

Equipment Replacements models can be grouped as: simple and complex ones. By simple models are meant those with a small number of unknown parameters. An instance, where only a small number of observations of time to fail are required to determine a near-optimal value of the critical age for preventive replacement may be an example for age-based replacement models (Baker and Scarf, 1995). The second group of models with a large number of parameters is characterized by high correlations between parameter estimates. This indicates that the available data is insufficient to distinguish between equally plausible parameter combinations (Scarf, 1997).

Recent discussions in econometric models are highlighted the question when to use a capital replacement modeling or an economic life modeling. By comparison of these two approaches, capital replacement modeling methods are evidently more application-oriented than second ones (Christer and Scarf, 1994; Scarf and Bouamra, 1995).

Pioneering approaches of equipment replacement studies are mainly addressed to the replacement of single machines or systems. Developed methods stated to multi-machine systems have mostly assumed linear production flows, with limited operational flexibility. A multi-period replacement model for flexible automated systems was developed by Lotfi and Suresh (1994). Their model was formulated as a nonlinear integer programming problem and was intended to serve as an analytical approximation along with closed queuing networks.

Obviously, it would be possible to mention more similar works from different authors on that topic, but this was comprehensively provided, for instance, by Fine and Freund (1990) or Cheevaprawatdomrong and Smith (2003).

MODELS DEVELOPMENT

The further presented equipment replacement decisions models are econometric-based methods. Econometric methods are in generally concerned with using relevant data for modeling relations between economic and business variables. In these methods one problem is the fact that the selection of variables is somewhat subjective. Their role in decision support for the equipment management and replacement consists of finding the adequate moment to change machine-tool in use or its part(s), based on a specified criterion. In the next subparagraphs several mathematical methods regarding the equipment replacement decision are described.

The OEC and ORC Dependence Based Method

The problem is to choose an optimal replacement policy such that sum of operating equipment cost (OEC) and replacement equipment cost (REC) per unit time is minimized.

In general, the calculation of operating costs (OC) requires the examination of various influencing parameters. Moreover, there is some difference of opinion about whether the wages of equipment operators should be included in the operating equipment cost (Sears et al 2008). In this method the wages are included to this cost. Because we are looking at all costs from cash flow perspective equipment, thus a replacement cost in our approach deals with present value analysis.

The instrumental assumptions of this method imply that at the beginning of every year, data regarding operation and replacement costs of a certain machine are collected. The data usually shows an increase in the operation cost, because of the damages in certain components of the machine. Some of these components may be replaced, thus the equipment operating cost are reduced. The replacement thereof implies costs with the materials and salaries and, hence, such costs have to be compensated through the savings which may be obtained pursuant to the reduction of operating costs. Thus, we want to determine an optimal replacement policy, able to minimize the sum of operating and replacement expenses during the period between two successive data collections.

Let us consider $c(t)$, the operating cost per time unit at the moment t, after replacement and c_r, the cost of a replacement. Then, the relation between the operating cost, replacement cost and time is shown in Figure 1.

The replacement policy is presented in Figure 2, with the following notations: $[0,T]$, the time interval regarding the collection of data on the machine and t_r, intervals when n replacements shall occur.

The assumed goal is the determination of the optimal interval between successive replacements, so that the sum of the operating and replacement cost $C(t_r)$ is minimal.

Then, $C(t_r)$ present the replacement cost during the period $[0,T]$, plus operating cost during the period $[0,T]$.

Figure 1. Relation between operation cost and replacement cost

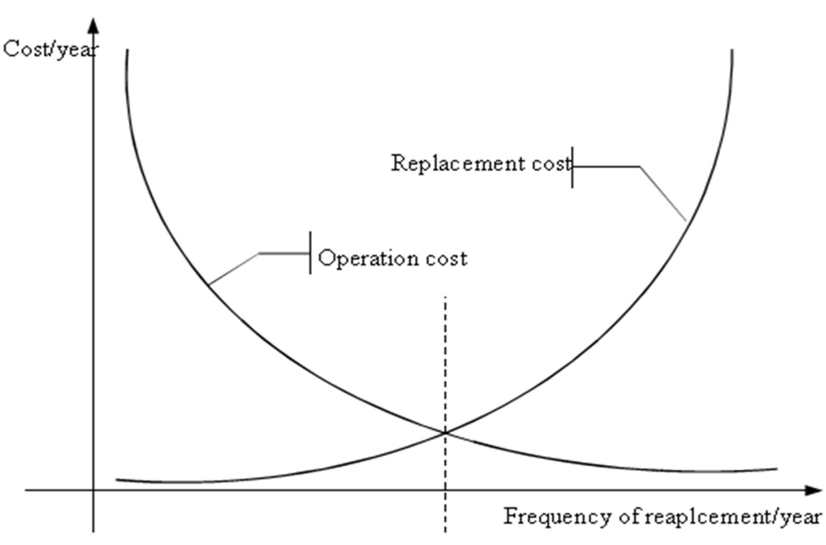
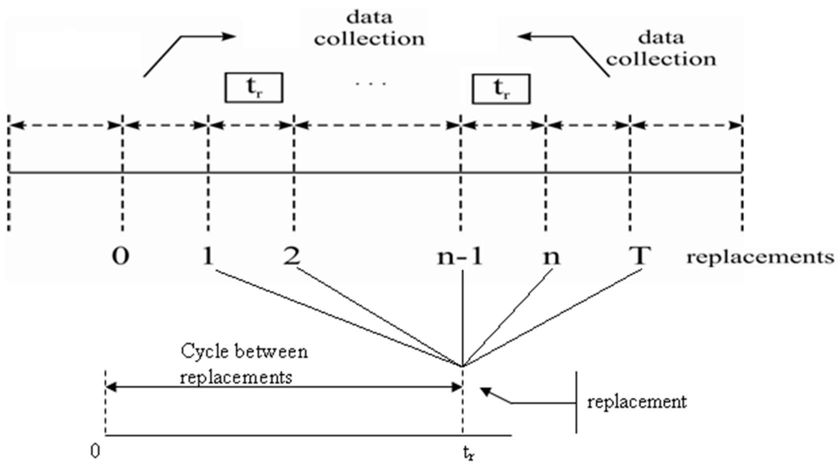


Figure 2. Graphical representation of the replacement policy



Replacement cost by period $[0,T]$ is calculated by

$$\sum C_r = n \cdot C_r \tag{1}$$

Thus, n are the number of replacements by period $[0,T]$ and C_r, the cost of one replacement.

The total cost per time unit $C(t_r)$, for the replacement performed at the moment t_r is: $C(t_r)$ i.e. total cost in the interval $(0,t_r)$ related to the length of the interval.

The total cost in the interval $(0,t_r)$ is the operating cost plus replacement cost.

$$C\left(t_r\right) = \int_0^{t_r} c\left(t\right) dt + C_r \approx \left[\frac{\int_0^{t_r} c\left(t\right) dt + C_r}{t_r}\right] \tag{2}$$

As we may see, the two different cost calculation procedures are similar, because the minimization of $C(t_r)$ is desired, depending on t_r.

Neither of the two procedures considers the time T_r (see Figure 3) required for performing a replacement.

If the time required for performing a replacement is considered, Equation 2 becomes:

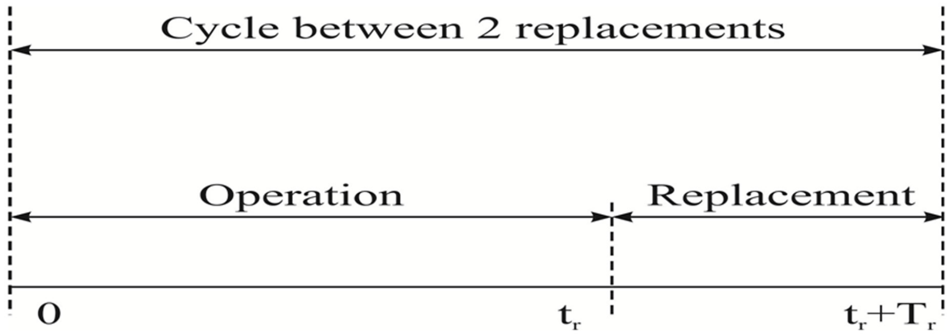
$$C\left(t_r\right) = \frac{\int_0^{t_r} c\left(t\right) dt + C_r}{t_r + T_r} \tag{3}$$

Even though this method is also applicable for job shop problems, the given procedures meet essential requirements of the Total Productive Maintenance concept that have a very important role to effective use the automated production systems like FMC and flexible manufacturing systems.

The Method of Replacing the Equipment at a Certain Age

We consider that the machine shall be used for a certain number of years. Further, we assume that a certain machine fabricates certain products, according to the production plan. In this case the goal, in order to minimize the total operating and replacement cost for a fixed time period, consists in the determination of the replacement policy establishing whether: at a certain age of

Figure 3. Time structure of cycle between two replacements



the machine; the latter should be replaced or left to operate continuously.

Let us use *I* to denote the age of the equipment (from the last replacement, with *n* plan periods of proper operation, until the end of the production plan); *c(a)* to represent the cost of operating the equipment for a plan period, when the equipment has age *a*; *J* to represent the age of the equipment from the moment of the last replacement, having (*n-1*) operating time periods until the end of the production plan; C_r, replacement cost; *C(I,J)*, total cost during the period when the equipment develops from age *I* to age *J*. The proposed goal consists in the determination of a replacement policy, so that the cost of operating and replacing the machine $C_n(i)$, along the following *n* time periods is minimal. When $C_n(i)$ has a minimal value, the smallest cost is defined as $f_n(i)$.

Ten weeks before the end of the production plan, two decisions may be made: continuous use or replacement of the machine. If it is decided that the machine should operate further, the equipment shall have age *4* when a new decision may be made (Figure 4).

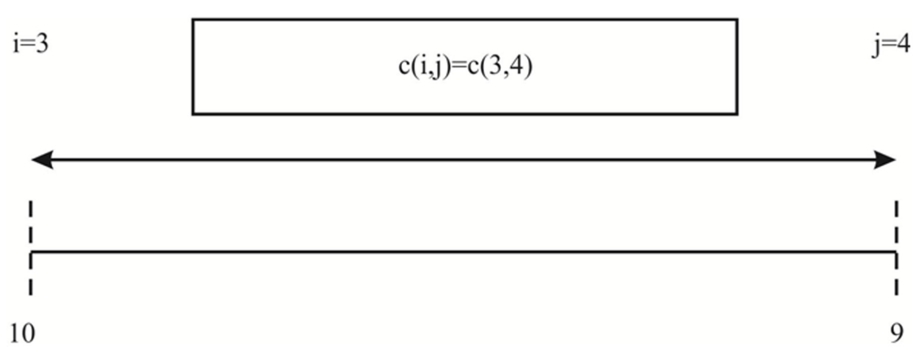
The total operating cost for the period (10, 9) is:

$$C\left(3,4\right)=C\left(3\right) \tag{4}$$

If the decision to replace the machine is made, then the total cost for the period (10, 9) shall be:

$$C\left(3,1\right)=C_r+C\left(0\right) \tag{5}$$

Figure 4. Replacement policy for the machine with age "4"



Thus, C_r is the replacement cost, and *C(0)*, the operating cost for a period, when the machine has age *0*.

The optimal replacement policy is graphically presented in Figure 5.

The mathematical model used for identifying this optimal policy has the following form:

Consider: $f_n(i)$, the minimal cost resulting from taking the best decision at the beginning of the period n plus the cost of the best decision taken on the remaining periods (*n-1*); *C(i,j)* represent the cost resulting from taking the decision at the beginning of the period *n*; $f_{n-1}(j)$, minimal cost by periods (*n-1*) remaining at the moment when the machine has the age *J*.

The cost by the *n* periods is:

$$C\left(i,j\right)+f_{n-1}\left(j\right), \tag{6}$$

so:

$$f_n\left(i\right)=\min\left[C\left(i,j\right)+f_{n-1}\left(j\right)\right] \tag{7}$$

with:

$$f_0\left(i\right)=0$$
$$j=i+1 \quad \text{or } 0$$

The equation (7) that may be solved by means of dynamic programming can be used to determine the replacement policy under a given specifications formulated in a section 'The problem statement'.

The Method of Replacement Based on the Existence of Equipment in Standby

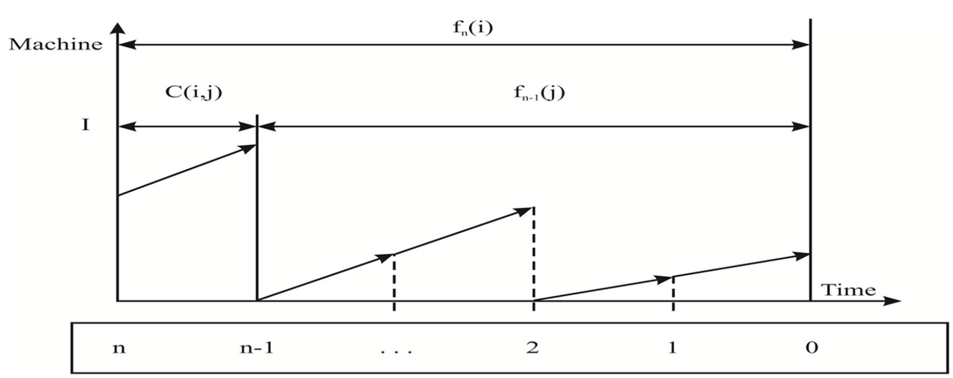
It implies the replacement of assets provided that the manufacturing flow contains a spare asset, and the operating cost increases with the use of the asset existing in production.

In this case, an optimal replacement policy must be determined, combined for the two assets, which shall minimize the total replacement and operating cost for a fixed time period.

The state of the production system at the beginning of a period shall be noted with *I*, where *I* is equivalent to the pair of numbers (*x, y*), where *x* refers to the asset (*A* or *B*) which is generally used, and *y*, to the age of the asset.

Consider: $C_x(y)$, the operating cost for a period; *j*, the state of the production system at the end of a period, where *j* is equivalent to (*x,y*); C_r = replacement cost, considered equal for both assets *C(i,j,)*, total cost of the system between the states of the system *i* and *j*. The time required for the replacement of an asset is a period when the replacement decision is made, and then the *stand by* asset becomes

Figure 5. Optimal replacement policy



operative. The proposed goal is the determination of an optimal combined policy for replacement/operation, so that the operating and replacement cost for the following n time periods is minimal. Figure 6 shows such a policy, where $n=10$, the system is in state $I=(B,2)$. At the beginning of period *10*, a decision is made to go on with asset B.

At the beginning of period *9*, a decision is made to replace asset *B* etc. The total minimal cost for replacement and operation, for the n periods is $f_n(i)$.

The cost of the first decision taken at the beginning of the period n is $C(i,j)$. At the end of this period, the system is in state j, having $(n-1)$ operating periods. Then the minimal cost for the remaining period $f_{n-1}(j)$ is:

$$Total\cos t = C\left(i,j\right) + f_{n-1}\left(j\right) \quad (8)$$

and

$$f_n\left(i\right) = \min\left[C\left(i,j\right) + f_{n-1}\left(j\right)\right] \quad (9)$$

In the next section we will describe the replacing policy model based on considering technological changes.

The Method of Replacing the Equipments Based on Technological Improvement in Finished Time Horizon

Considers that the replacement of an old machine by a new one not always is an exact copy of the old one, but that the latter is better, so that operating and maintenance costs are smaller, efficiency is higher etc. The following model aims at determining the way how the new available machines may be used with a successful purpose, considering that the time period is fixed and finite.

Consider: n, the number of operating periods (periods when the machine must operate); $C_{p,i}$, maintenance cost of the current equipment in the period I ($i = l,2,...n$); $S_{p,i}$, sales value of the current equipment at the end of the time period; A, purchase cost for the new, better equipment; $C_{t,j}$, maintenance cost of the new machine in the dj period after installation ($j=l,2,...n$); $S_{t,j}$, sale value of the new equipment at the end of the operating period j; r – update factor.

The method aims at determining the value T when replacement should be made with the new, better machine (Figure 7), $T = 0, 1,2,....,n$.

Figure 6. Replacement policy given the existence of the standby asset

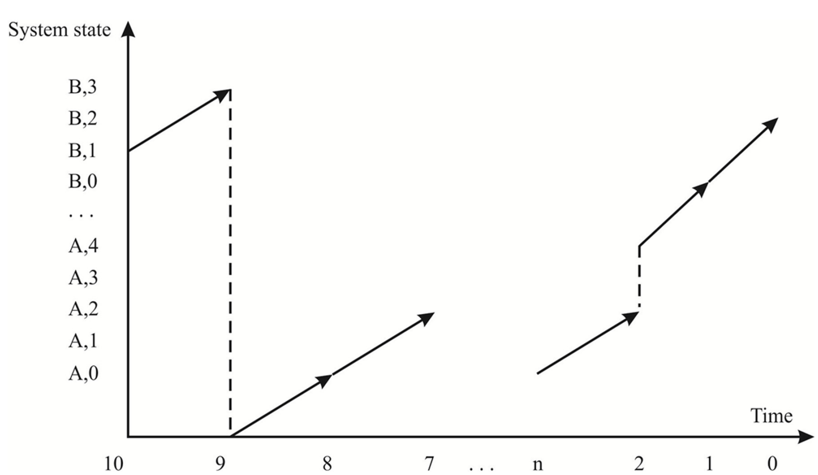
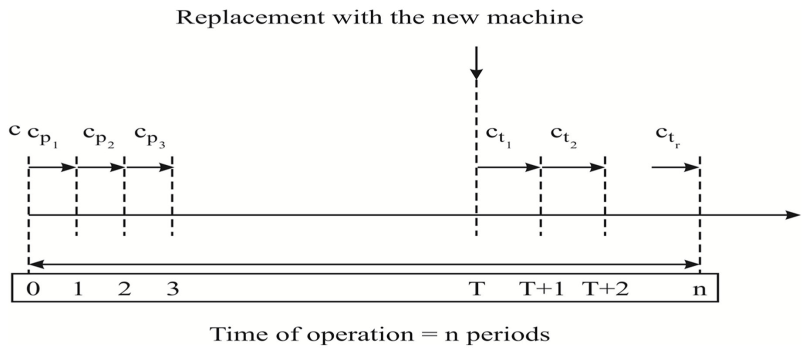


Figure 7. Graphical calculation of T value



The total updated cost for the *n* periods when replacement occurs at the end of the *T* period is: *C(T)*, updated cost for the maintenance of the current machine in the period *(0,T)*, plus the updated maintenance cost of the new machine in the period *(T,n)*, plus the updated purchase cost of the new machine, minus the updated sale value of the current equipment at the end of the *T* time period, minus the updated sale value of the new equipment at the end of period *n*:

$$C(T) = \left(C_{p,1} \cdot r^1 + C_{p,2} \cdot r^2 + \ldots + C_{p,T} \cdot r^T\right) + \left(C_{t,1} \cdot r^{T+1} + C_{t,2} \cdot r^{T+2} + \ldots + C_{t,n-T} \cdot r^n\right) + A \cdot r^T - \left(S_{p,T} \cdot r^T + C_{t,n-T} \cdot r^n\right) \tag{10}$$

Hence,

$$C\left(T\right) = \sum_{i=1}^{T} C_{p,i} \cdot r^i + \sum_{j=1}^{n-T} C_{t,j} \cdot r^{T+j} + A \cdot r^T - \left(S_{p,T} \cdot r^T + C_{t,n-T} \cdot r^n\right) \tag{11}$$

As the sole unknown variable is *T*, the minimisation of *C(T)* does not raise further related issues.

DISCUSSION AND CONCLUSION

Because of the fact that above presented methods are more or less applicable based on specific theoretical preconditions, it can be handy to view the given problem from a wider user base. Prior to analysing decisions about equipment replacement in flexible manufacturing cells from practical point of view it is useful recognize two different approaches: deterministic or probabilistic. The probabilistic decisions to replace machines referred to as preventive actions are those decisions where the risk is given by the impossibility to exactly determine the moment when such machine falls or the transition moment from proper operating state to non-operating state. Another source of risk is given by the impossibility to determine the state of the equipment when no inspection or other maintenance activity occurs. Let us consider that there are only two states of the equipment that are always known: a proper operating state or a non-operating state. Then, in order to avoid equipment stoppages in flexible manufacturing cells, the positive replacement decision should come during a proper operating state and accordingly should have a preventive character. In such a way understood preventive replacement for fixed assets implies two conditions:

- *the total* replacement *cost shall be higher after the fall itself at the moment when the preventive replacement is made;*
- the replacement of the machine before the fall itself does not affect the chance that the equipment may fall at the following moment.

Therefore, preventive replacement is only justified when the rate of replacement grows. In case the machine is damaged, specialists in the department should increase the preventive replacement activities. This may lead to a mistake, as the preventive replacement of machines or their parts is not always justified.

Another above presented approach to equipment replacement decisions is based on considering technological shifts. Even though it is often assumed that an acceleration in technological improvement should result in a more rapid introduction of new technology, according to Cheevaprawatdomrong and Smith (2003), rapid technological improvement may not and indeed should not necessarily lead to more rapid replacement of old technology. In addition, as regards to methods of replacing the equipments based on technological improvement, it has to be 'calculated' with known difficulties and problems such as:

- workers' resistance, as they are used to the old machine;
- lack of will to change the work style;
- fear of the unknown, i.e. workers are scared that they will lose their jobs pursuant to the introduction of new technologies or that they won't be able to adjust to the new working requirements;
- lack of support regarding specialized documentation;
- difference of opinions regarding the operation of the equipment.

Accordingly, a management attitude toward new manufacturing technology will play a major role in determining whether a firm will acquire such technology (Dorf and Kusiak,1994).

REFERENCES

Baker, R. D., & Scarf, P. A. (1995). Can models fitted to small data samples lead to maintenance policies with near-optimum cost? *I.M.A. Journal of Mathematics Applied in Business and Industry*, *6*, 3–12.

Bean, J. C., Lohmann, J. R., & Smith, R. L. (1994). Equipment replacement under technological change. *Nav Res Log*, *41*, 117–128. doi:10.1002/1520-6750(199402)41:1<117::AID-NAV3220410108>3.0.CO;2-U

Cheevaprawatdomrong, T., & Smith, R. L. (2003). A paradox in equipment replacement under technological improvement. *Operations Research Letters*, *31*, 77–82. doi:10.1016/S0167-6377(02)00153-0

Christer, A. H., & Scarf, P. A. (1994). A robust replacement model with application to medical equipment. *The Journal of the Operational Research Society*, *45*, 261–275.

Dehayem Nodem, F. I., Kenné, J. P., & Gharbi, A. (2009). Hierarchical decision making in production and repair/replacement planning with imperfect repairs under uncertainties. *European Journal of Operational Research*, *198*(1), 173–189. doi:10.1016/j.ejor.2008.09.001

Dorf, R. C., & Kusiak, A. (1994). *Handbook of design manufacturing and automation*. John Willey & Sons Inc. doi:10.1002/9780470172452

Fine, C., & Freund, R. (1990). Optimal investment in product-flexible manufacturing capacity. *Management Science*, *36*, 449–466. doi:10.1287/mnsc.36.4.449

Flynn, J., & Chung, C. (2004). A heuristic algorithm for determining replacement policies in consecutive k-out-of-n systems. *Computers & Operations Research*, *31*(8), 1335–1348. doi:10.1016/S0305-0548(03)00094-7

Freeman, J. (1996). Analyzing equipment failure rates. *International Journal of Quality & Reliability Management*, *13*(4), 39–49. doi:10.1108/02656719610114380

Groover, M. P. (2001). *Automation, production systems, and computer integrated manufacturing*. New Jersey: Prentice Hall.

Hartman, J. C., & Ban, J. (2002). The series-parallel replacement problem. *Robotics and Computer-integrated Manufacturing*, *18*, 215–221. doi:10.1016/S0736-5845(02)00012-1

Hritonenko, N., & Yatsenko, Y. (2007). Optimal equipment replacement without paradoxes: A continuous analysis. *Operations Research Letters*, *35*(2), 245–250. doi:10.1016/j.orl.2006.03.001

Jianhua, S., & Fujimoto, Y. (2003). Dynamic scheduling of large-scale flow shops based on relative priority approach. *IEE J Trans. EIS*, *123*(10), 1858–1864. doi:10.1541/ieejeiss.123.1858

Korbaa, O., Camus, H., & Gentina, J.-C. (1997). Cyclic scheduling with overlapping production cycles. *Proceedings of 18th International Conference on Application and Theory of Petri Nets*, Toulouse, France, June 1997.

Leung, L. C., & Tanchoco, J. M. A. (1990). Multiple machine replacement analysis. *Engineering Costs and Production Economics*, *20*, 265–275. doi:10.1016/0167-188X(90)90074-R

Lotfi, V., & Suresh, N. C. (1994). Flexible automation investments: A problem formulation and solution procedure. *European Journal of Operational Research*, *79*(3), 383–404. doi:10.1016/0377-2217(94)90054-X

Love, C., Zhang, Z., Zitron, M., & Guo, R. (2000). Theory and methodology: A discrete semi-Markov decision model to determine the optimal repair/replacement policy under general repairs. *European Journal of Operational Research*, *125*(2), 398–409. doi:10.1016/S0377-2217(99)00009-0

Sarache Castro, W. A., Castrillón, O. D., Gonzales, G., & Viveros Folleco, A. (2009). A multi-criteria application for an equipment replacement decision. *Ing. Desarro., 25,* 80-98. ISSN 0122-3461

Scarf, P. A. (1997). On the application of mathematical models in maintenance. *European Journal of Operational Research*, *99*, 493–506. doi:10.1016/S0377-2217(96)00316-5

Scarf, P. A., & Bouamra, O. (1995). On the application of a capital replacement model for a mixed fleet. *I.M.A. Journal of Mathematics Applied in Business and Industry*, *6*, 39–52.

Sears, S. K., Sears, G. A., & Clough, R. C. (2008). *Construction project management: A practical guide to field construction management*. New York, NY: John Wiley.

Stinson, J. P., & Khumawala, B. M. (1987). The replacement of machines in a serially dependent multi-machine production system. *International Journal of Production Research*, *25*, 677–688. doi:10.1080/00207548708919870

Suresh, N. C. (1991). An extended multi-objective replacement model for flexible automation investments. *International Journal of Production Research*, *29*, 1823–1844. doi:10.1080/00207549108948052

Suresh, N. C. (1992). A generalized multimachine replacement model for flexible automation investments. *IIE Transactions*, *24*, 131–143. doi:10.1080/07408179208964210

Tanchoco, J. M. A., & Leung, L. C. (1987). An input–output model for equipment replacement decisions. *Engineering Costs and Production Economics*, *11*, 69–78.

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 401-411, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 27
Enhancing Engineering Education Learning Outcomes Using Project-Based Learning:
A Case Study

Mousumi Debnath
Jaipur Engineering College and Research Centre, India

Mukeshwar Pandey
Jaipur Engineering College and Research Centre, India

ABSTRACT

This paper presents a case study of how project-based learning (PBL) can be seen as a pedagogical innovation for Jaipur Engineering College and Research Centre, Jaipur, India (JECRC) for better recruitment drives for on-campus recruitments. The practical knowledge of engineering, basic knowledge of engineering design, soft skills or personal competences can be correlated with the academic performance and recruitment status. Project based learning (PBL) is a learner-centric pedagogy where the learner is expected to take responsibility for his or her own learning. PBL uses in-depth and rigorous classroom projects to facilitate learning and assess student competence. Students have projects as a compulsory course in their curriculum in the final semester of engineering. The challenge to acquire knowledge and skills during their project gives a student an opportunity to develop their weaker skills and enhance their practical knowledge of engineering. This study has been successful in helping students acquire a high rate of actual skill and technical learning. The learning outcomes of the PBL-course can be correlated with their success in recruitment and academic performances.

DOI: 10.4018/978-1-4666-1945-6.ch027

INTRODUCTION

Today's students need to see and understand the relevance, reality, and authenticity of concepts and apply them to their lives in the real world. One unique teaching method that is proving to be effective in the classroom is project-based learning (PBL). PBL is a hands-on approach to learning. It incorporates a number of need-to-know concepts that they must use technology and inquiry to respond to a complex issue, problem or challenge, understand, and apply in order to complete the project. Project-based education is a learning environment congruent with the principles of student-and competence-centred vision. This has been an ongoing innovation since last 40 years. It can be seen as a pedagogical method which integrates theory and practice by means of problem solving of working life issues (Laynea *et al.,* 2008).

Bransford and Stein (1993) have defined PBL as a comprehensive instructional approach to engage students in sustained, cooperative investigation. The PBL approach is appropriate to acquire generic skills such as problem-solving, communication and teamwork (Wolfs *et al.*, 1997). An important piece of PBL is incorporating technology into projects. The team projects in PBL have a contextual focus enabling students to understand why they are learning the particular content and how it will be applied in the 'real world'.

PBL goes beyond generating student interest. Well-designed projects encourage active inquiry and higher-level thinking (Thomas, 1998). The students' major challenge is to acquire new understanding. PBL helps to enhance creative thinking skills by showing that there are many ways to solve a problem when they are connected to problem-solving activities. The students are helped to understand why, when, and how those facts and skills are relevant (Bransford *et al.,* 2000). Within the project based learning framework students collaborate, work together, and take responsibility for their own learning.

CHALLENGES OF PROJECT BASED LEARNING

Because of an increasing quality concern for higher education, additional attention is being paid to new educational principles with a more student- and competence-centred vision (Van der Bergh *et al.,* 2006). Project-based learning is one of the learning environments congruent with these principles (Van de Bergh *et al*., 2006). Project-based learning offers a wide range of benefits to both students and teachers. Academic research supports the use of project-based learning in college/school to engage students, cut absenteeism, boost cooperative learning skills, and improve academic performance. The major challenges of PBL include enhanced student participation in the learning process (active learning and self-learning), enhanced communication skills, addressing of a wider set of learning styles, and promotion of critical and proactive thinking and finally making them more competent to get jobs in on campus recruitment drives. The real-world problems capture students' interest and provoke serious thinking and motivate them for self-learning process. PBL also facilitates the development of many of the "soft skills" demanded from engineering graduates (Hadim & Esche, 2002). Soft skills and "generic skills" are interchangeable phrases in terms of the categorization of non-technical skills. For students, project-based learning helps to overcome all challenges and convert them into their major success. They include:

- Increased attendance, growth in self-reliance, and improved attitudes toward learning (Thomas, 2000).
- Academic gains equal to or better than those generated by other models, with students involved in projects taking greater responsibility for their own learning than during more traditional classroom activities (Boaler, 1999; SRI International, 2000).

- Opportunities to develop complex skills, such as higher-order thinking, problem-solving, collaborating, and communicating (SRI International, 2000).
- Access to a broader range of learning opportunities in the classroom, providing a strategy for engaging culturally diverse learners (Railsback, 2002).

Program assessment is a vital activity in order to determine the outcomes of student engagement in PBL. These outcomes include programme issues, knowledge, skills, attitudes and identity and post educational professional performance. This is shown in Figure 1. Knowledge and skills are, generally, easy to directly measure by common assessments used in nearly all engineering courses; usually conducted via graded reports, presentation, and other student work. Attitudes are often hard to measure, particularly within the timeframe of a single course where changes often manifest later upon self-reflection. General use of these assessment tools to measure the skills, outcomes and competencies developed from participation in PBL requires further evaluation. The major challenge and the final outcome in the professional programms/course are to obtain jobs (Swan *et al.,* 2009). Recruitment is seen as the final challenge where all the attributes are taken into consideration. PBL plays an important role in making a student technically strong and competent to face interviews.

The research question addressed in this study is the suitability of PBL in achieving the desired learning outcomes like practical knowledge of engineering, basic knowledge on engineering design, soft skills or personal competences and its impact on the student's performance in on-campus recruitment drive. So, a proposed model to correlate PBL with academic proficiency and recruitment would work upon.

Figure 1. Outcomes of project based learning

It was assumed that it depends on two major hypotheses. The first is to believe that the academic performance is not always associated with skill and technical expertise. Thereby, PBL learning delivery requires providers to alter their in-class approach and demands students to be more independent learners. The second concept is the students who are innovative and sincere in their project work have an interdisciplinary learning approach. Thus due to this interdisciplinary approach, these students are well placed in the on-campus recruitment drives. Thus by systematic monitoring and evaluation of the various aspects of the recruitment status of the institute and the performance of the students in PBL, the self-evaluation quality assessment of the institute was performed.

In the first part different kinds of evaluation methods have been compared, whereas the second one refers to specific, self-evaluation quality assessment research which has been done at Lublin University of Technology, Poland (Wac-Włodarczyk & Billewicz, 2008). Its main objective was to conduct precise number of surveys between different groups of people.

CASE STUDY: PROJECT BASED LEARNING (PBL) AT JECRC

For the course of study, the 444 final year students of Jaipur Engineering College and research centre (JECRC) of 2010 enrolled in project based learning course/s were considered and critically assessed. JECRC is one of the leading technical institutions in Northern India. The institute offers a four-year Bachelor of engineering degree in Electronics and Communication Engineering

(ECE), Electrical Engineering (EE), Computer Science and Engineering (CS), Biotechnology Engineering (BT), Information Technology (IT) and Mechanical Engineering (ME). All the courses are approved by All India Council for Technical Education (AICTE) at national level and affiliated to Rajasthan Technical University (RTU), Kota in the state. The enrollments are strictly on the basis of ranking in the various competitive examinations; such as All India Engineering Entrance Examination (AIEEE) and Rajasthan Pre-Engineering Exam (RPET). The institute nurtures the essence of growth in education and its holistic approach focuses on the overall development of its students. As a part of engineering curricula all the students in engineering programs/courses under the RTU syllabus have to pursue a definitive project. The duration of the project-based learning varies from 45 days to six months. The following sections discuss a few examples of the projects undertaken as PBL and then finally analyse the impact of PBL on academics and recruitment.

Challenging Examples of PBL at JECRC

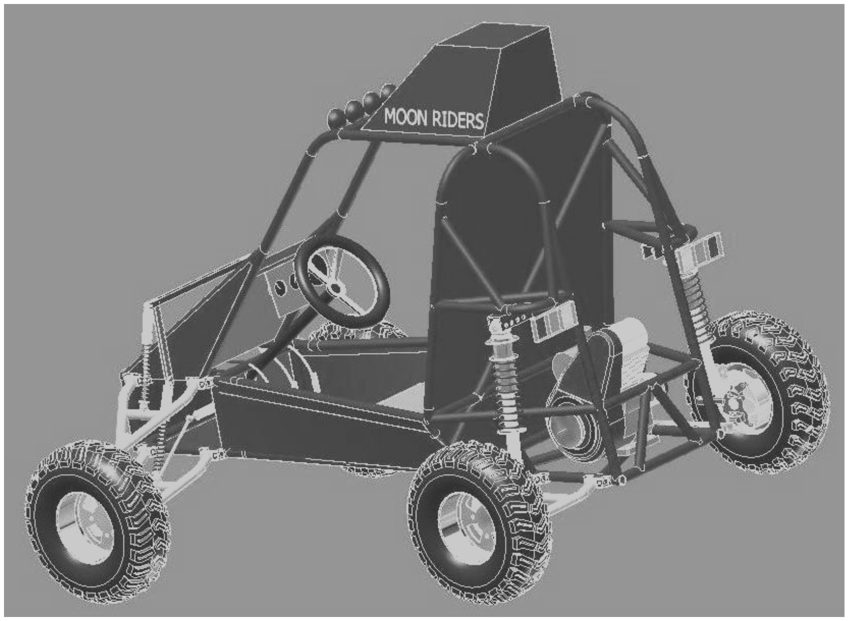
The project-based learning at JECRC depicted various models of instruction such as "active learning", "contextual learning", "design-based modelling", "collaborative learning", "technology-based education" and "design experiments". A number of experimental and theoretical investigations have been done on the application, development, and benefits of project-based learning in education. It was found that the role of the student changes from that of a consumer in a traditional classroom based setup to a more involved role of an active learner in project-based learning (Eugene, 2006).

All the PBL projects at JECRC were asked to be built around thematic units or the intersection of topics from two or more disciplines. For example: a 6 months project on 'Fire vigilance robot" was submitted by students of electronics and IT. In this project an implementation of artificial intelligence and electronic circuitry helped in devising an autonomous machine capable of avoiding obstacles coming in its way. It has a 360 degrees movement, temperature sensing, raising alarms, raising alerts using buzzer alarms. Thus, a multidisciplinary project-based learning helped students to develop competency in engineering, science and technology skills, teamwork and management, and then giving them a competitive advantage in industry.

All these projects involved students in a constructive investigation. The investigations were goal directed process involving inquiry, knowledge building, and solution. All these investigations also involved design, decision-making, problem-finding, problem-solving, discovery and r model-building processes. In this context, a project by the mechanical students on automobile can be taken up as an example. The 19 member team under the leadership of final year mechanical engineering students at JECRC, conducted market surveys and carried out discussions on emerging trends in off road vehicles available across the globe. A reverse engineering paradigm was followed for designing a vehicle "Moon rider" (Figure 2). The final 3-D CAD model was modelled in Autodesk Inventor 2010. A PVC prototype was made of chassis to have an accurate idea about the various constraints and to have a real experience. To save time and cost for physical testing, the finite element analysis (FEA) was carried on ANSYS Workbench 10 and various load analysis was done.

A steel prototype was made after selecting weld type and material for chassis. Fabrication of chassis and the weld strength checks were made to ensure the accuracy of loading conditions. The powering engine of the vehicle and the important aggregates such as suspension and steering were calculated. The steering system was self-customised and customer centred to meet the needs and to add additional value. A customer oriented marketing plan was made in accordance to make a business model to manufacture 4000 vehicles

Figure 2. Moonriders: Innovation of mechanical engineering

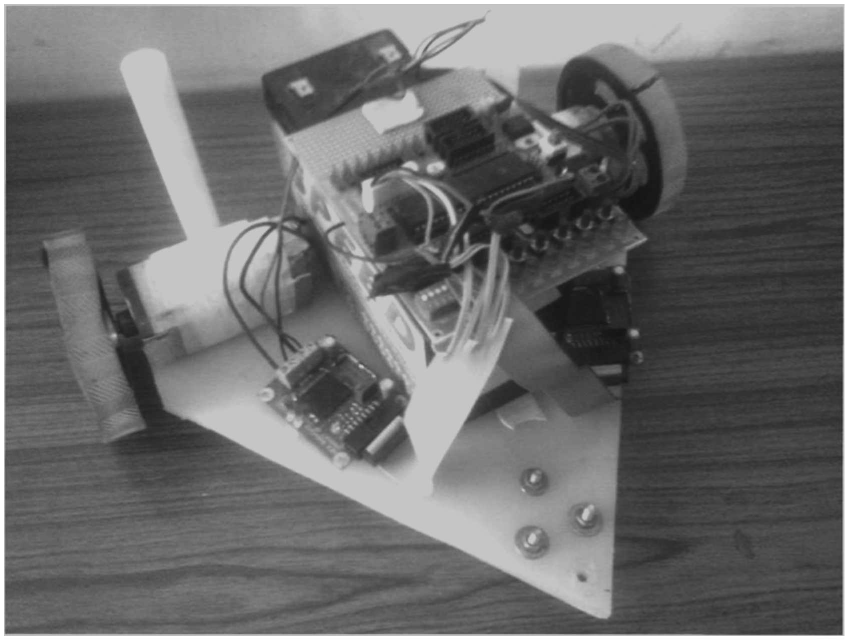


in a year. Proper industrial setup with plant layout was planned and the costing was calculated efficiently so as to minimise the manufacturing cost. After the final testing, the vehicle was transported for final evaluation where various static and dynamic evaluations on the vehicle were done. This project was highly appreciated in different state level and national level competitions and was presented in a national competition for automobile designs, namedBAJA SAEINDIA 2010.

However, in order to be considered a PBL project, the central activities of the project was asked to involve the transformation and construction of knowledge (by definition: new understandings, new skills) on the part of students. The mechanical engineering project could transform their classroom training into their practical innovation and helped them in acquiring the intended learning outcomes within knowledge and skills. This, in turn, also increased the quality of the education imparted by the institute and helped them to stand apart from the rest of the institutes. Another innovative project by student group from the department of computer engineering at JECRC was the **"**Twitli" *Twitter Client for Android* project. The objective of this project was to illustrate the requirements of Twitli, which is a Twitter client for Android phones (social networking website http://www.twitter.com). Twitli will facilitate users to access and use Twitter with all facilities provided by twitter.com. These smart phones use technologies like Android SDK, Android Components (Dalvik VM, Google API etc.) and Java with Android packages, XML etc., Twitter API wwwapiwiki.twitter.com and www.developer.android.com. This project helped the students to create a learning environment towards attaining a competence centred vision.

Projects embody characteristics that give them a feeling of authenticity to students. These characteristics can include the topic, the tasks, the roles that students play, the context within which the work of the project is carried out, the collaborators who work with students on the project, the products that are produced, the audience for the project's products, or the criteria by which the products or performances are evaluated. In this context one the project, named Trinity, by a group of final

Figure 3. Trinity: Innovation of the electronics and communication engineering

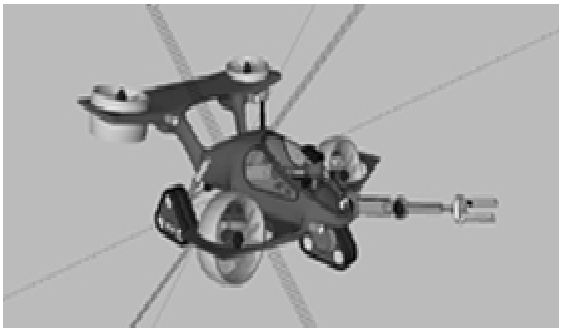


year electronics and communication engineering students can be considered as an example. Trinity project was made to design a control system to interact with a robotic platform wirelessly using voice, remote and visual communication technique was developed (Figure 3). In this robot a voice and remote based control has been completed but visual communication is yet to be implemented. This project also helped students to develop their competence towards learner centred pedagogy where students took full responsibility of making their ideas and innovation into a complete product.

The criteria for selection of a project are centrality, driving question, constructive investigations, autonomy, and realism. Students encounter and learn the central concepts of the discipline via the project. The central teaching strategy behind this PBL follows traditional instruction in such a way that the project serves to provide illustrations, examples, additional practice, or practical applications for material taught initially by other means. However, these "application" projects help students to learn things that are outside the curriculum ("enrichment" projects). In this context, final year students of electrical engineering made a promising project on mission-configurable robot that accommodates a wide range of sensors, including a dexterous manipulator and named it "ABHIMANYU". This robot can performs basic reconnaissance, increase situational awareness in dangerous scenarios for public safety professionals. It is both a terrestrial and an aerial vehicle as shown in Figure 4. This project helped students to make their creativity and constructive ability which gave a new dimension. While working on this project, students could put their ideas into realistic shape.

PBL incorporates real-life challenges where the focus is on authentic (not simulated) problems or questions and where solutions have the potential to be implemented. The Intelligent Notice Board is an initiative to design and implement an interactive notice board for use in college campus. The project is in the initial circuit design stage and is to be implemented using larger LCD screen. Initial control system built using Atmega-16, Max-232 and a 16*2 LCD. This project is a unique example of real life challenges.

Figure 4. Abhimanyu: Innovation of the electrical engineering



But it was finally concluded that the initial step for making a good project and learning from the project depends upon the first step of the project selection. If students can select projects themselves, the outcome is both interesting and exciting (Pucher *et al.*, 2003). Such a project can assure a quality improvement to the academic performance and personality development of the student.

Strategic Analysis of PBL in JECRC

A number of engineering schools have reframed their education around project-based learning (Beddos *et al.,* 2010) or active learning but recruitment is always the major motivator for professional study. In this study, the recruitment data from 2007-2010 was collected from the HR department of JECRC showing the number of students recruited in various engineering courses (Figure 5). On the basis of the recruitment data collected and analysed, it can be concluded that JECRC is blessed with high number of successful on-campus recruitment. This is clearly demonstrated in Figures 5 and 6. However, the question was how to increase the recruitment number in engineering departments other than CSE and ECE. There may also be several other factors apart from PBL that could have influenced the recruitment status of the institute.

There is some indication that PBL programs can help attract a more diverse population of students in engineering. The popularity of these experiences with students is clear, but it is not clear if this leads to any overall benefits to recruiting in engineering. There is virtually no quantitative assessment of the benefits of PBL experiences to professional trajectory. It is not fully clear if companies view this as a way to attract and retain qualified engineers, value the unique skills developed in engineers with these experiences, etc. Therefore, the impacts in this area require further study.

Figure 5. Recruitment statuses in JECRC

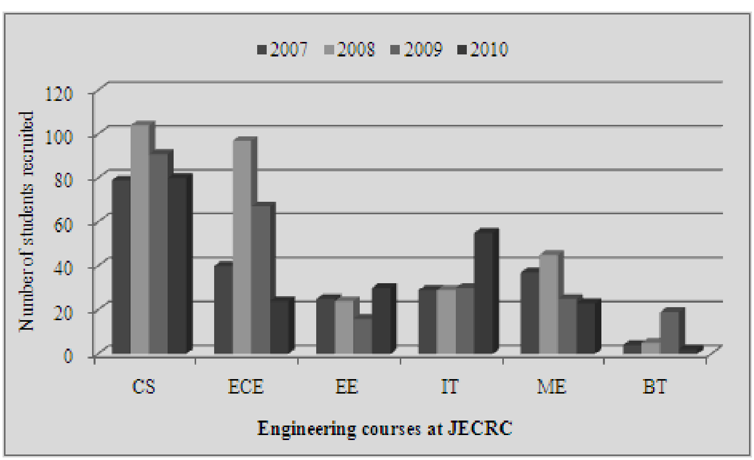
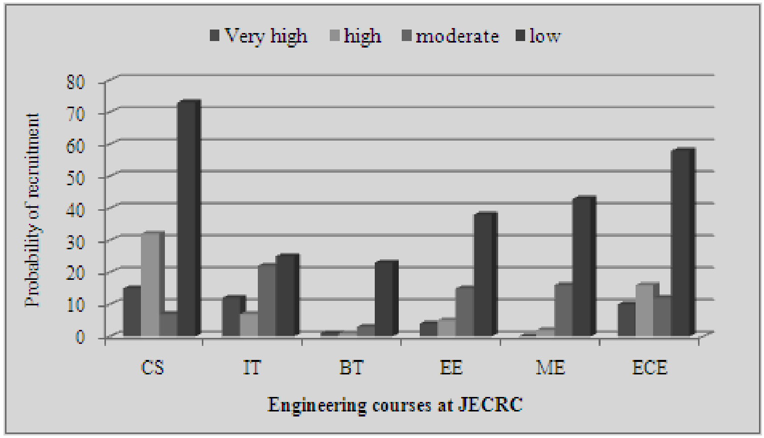


Figure 6. Probability of recruitment in engineering department of JECRC



An assessment of project based learning of the final year engineering students of the year 2010 at JECRC was done and correlated to the academic performance and also on campus recruitment of the students. The parameter taken were skills learnt during PBL, hands on training, innovation in the project, and practical application of the project. A grading scale for PBL was taken up from 10, 20, 30 and 40 for below average, average, good and excellent projects. The assessment grades were collected from the respective departments. Further it was found that students worked for these projects in groups as well as individually. The projects were completed in industries and as well as in the JECRC. The academic performance was also correlated taking into consideration their performance in school, high school and engineering studies and given a grading scale (below 60% =5; 60-69%=10; 70-79%=15; 80-89%=20; 90-100%=25). Students of JECRC have shown their abilities/potential to obtain jobs in top companies. They are usually placed in their final year while still studying in the JECRC In year 2010, 21 companies visited the campus for recruitments. To grade the recruitment, a scaling was done on the basis of the status of the company in terms of the salary package (excellent pay =15; moderate =10; average =5) and also the attempt the candidate made in securing the job (first attempt 21; second attempt 20 and so on).

On the basis of the aggregate scoring for academics, PBL and recruitment status of individual candidate, all students were further categorized for their very high (150-200), high (120-149), moderate (80-119) and low (below 80) probability for recruitment in the on campus placement drives as shown in Figure 6.

It was found that CS, IT and ECE had a very high probability of getting placed in good companies followed by ME and EE whereas BT had a low chance of recruitment. This was further investigated considering each variable (Table 1).

A relationship of PBL with recruitment was assessed. When the average of the variables were plotted, in almost all the departments, a direct relationship of an increase in the PBL performance could be seen with that of the recruitment status except for the BT, This is also shown in Figure 7.

It was found that in the year 2010, most of the on-campus recruitments were for IT and software based companies where BT students could not perform better in comparison to the other students. It was obvious that students with high academic performance would be good in PBL and a correlation between the academic performance and PBL was also analysed after plotting the mean of

Table 1. Data analysis of the engineering departments on the basis of academics, PBL and recruitment status in JECRC

Departments	Academics	PBL*	Recruitment status	Probability of on campus Recruitment
CS	51.33	30.00	86.86	Very High
	46.90	29.14	54.68	High
	45.00	25.71	23.71	Moderate
	41.36	28.00	0.00	Low
IT	51.25	25.00	88.16	Very High
	46.42	27.14	62.57	High
	41.36	20.45	29.13	Moderate
	39.20	23.60	0.00	Low
ECE	53.50	29.00	87.80	Very High
	52.18	27.50	55.87	High
	43.88	26.66	33.55	Moderate
	39.83	26.72	0.00	Low
ME	50.00	30.00	46.00	High
	41.94	25.55	19.82	Moderate
	33.47	26.58	0.00	Low
EE	55.00	37.50	85.25	Very High
	46.66	26.66	55.00	High
	40.00	28.23	20.23	Moderate
	32.50	23.68	0.00	Low
BT	50.00	40.00	61.00	Very High
	50.00	30.00	61.00	High
	53.33	40.00	0.00	Moderate
	41.50	21.00	0.00	Low

* Project Based Learning
CS = Computer Science and Engineering
IT = Information Technology
EE = Electrical Engineering
BT = Biotechnology Engineering
ECE = Electronics and Communication Engineering

the probabilities (Figure 8). Except in EE and BT, there was no direct correlation between the variables. It was revealed that only in the BT and EE disciplines, those students who were academically good were research oriented and showed better performance in PBL. Most of the departments except BT, worked on PBL in teams ranging from 2 to 6. The team spirit, networking along with on-site hands-on guidance for building practical skills helped the students to gain all the qualities required to be hired.

CONCLUSION

PBL emphasizes on students' learning in cooperation with others, and on their own participation in the act of learning. From the study, it was concluded that the probability of recruitment increases for those students who are innovative, skilful, disciplined and dedicated to PBL. The fundamental knowledge of engineering, engineering design, soft skills or interpersonal competences can be correlated with the academic performance and recruitment status. These students acquire the skills and competencies during PBL and this experience help them to secure a job in the on-campus recruitment drive. Hence it can be concluded that PBL is one of the major factors related to performance in on-campus recruitment at JECRC.

Thus, such an educational system is allowing students to obtain an acknowledged degree in an academic education, acquire specialised hands-on-training at the same time helping gain confidence for their recruitment programmes for their professional growth. It stresses the social dimension of learning, its negotiated nature, the need for a process in which learning is "anchored" in the real world. From the onset, it was realized that offering a PBL course such as project dissertations/training would require different teaching and learning strategies. Students and teaching faculty also acknowledged the same in the study conclusion.

It's important to note that throughout the project, the students have to employ their innovative skills and improvise the technique. During PBL, the students have to learn to move from commonsense ideas to more substantiated ideas based on practical experience and theoretical reflection, after all, that's what engineering is all about. A number of items need to be evaluated with respect

Figure 7. Relationship of recruitment and PBL in different engineering departments in JECRC

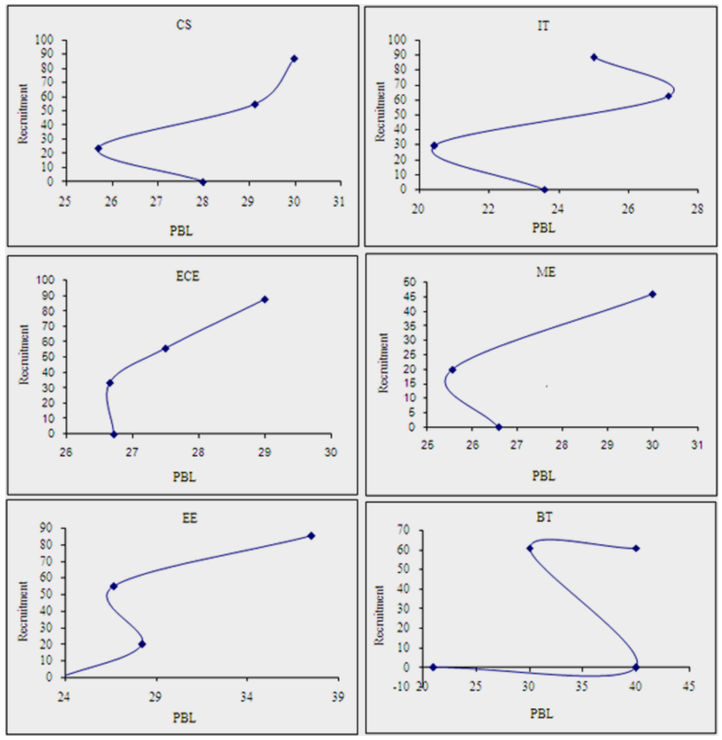
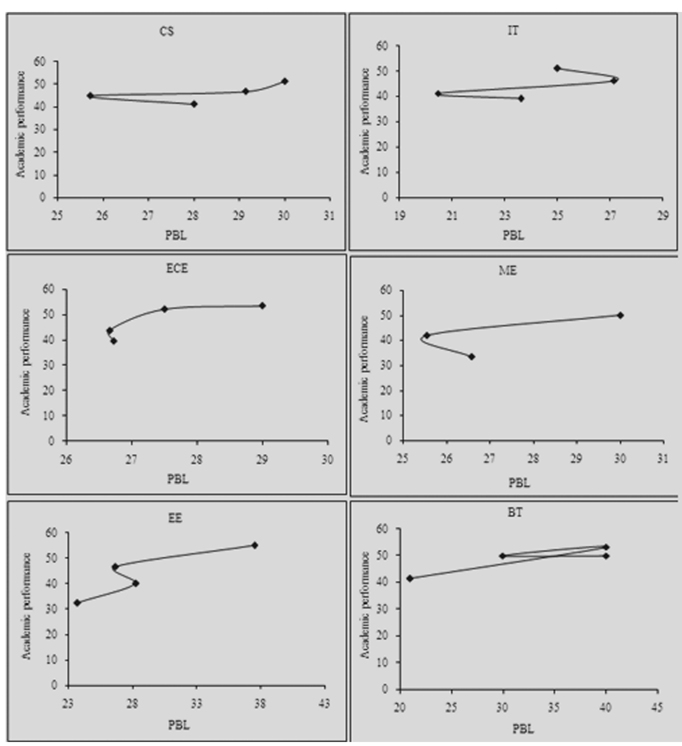


Figure 8. Relationship of academic performance and PBL in different engineering departments in JECRC



to the skills that the students develop by their involvement in PBL efforts. This study voiced the importance of PBL for learners in a quality higher education, the results of which will be recognized by society. PBL-based learning creates opportunities for groups of students to investigate meaningful questions and think critically. Answers to these questions can provide evidence that PBL is of value for a sustainable engineering education across a career.

REFERENCES

Beddoes, K. D., Jesiek, B. K., & Borrego, M. (2010). Identifying opportunities for collaborations in international engineering education research on problem based and project based learning. *Interdisciplinary Journal of Problem-Based Learning*, *2*(2), 7–34.

Boaler, J. (1999). Mathematics for the moment, or the millennium? *Education Week*.

Bransford, J. D., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school* (pp. 206–230). Washington, DC: National Academy Press.

Bransford, J. D., & Stein, B. S. (1993). *The ideal problem solver: A guide for improving thinking, learning, and creativity* (2nd ed.). New York, NY: Freeman & Co.

Eugene, C. (2006). How to teach at the university level through an active learning approach? Consequences for teaching basic electrical measurements. *Measurement*, *39*, 936–946. doi:10.1016/j.measurement.2006.09.001

Hadim, H. A., & Esche, S. K. (2002). Enhancing the engineering curriculum through project-based learning. In *Proceedings of the 32nd Annual Frontiers in Education Conference*.

International, S. R. I. (2000). *Silicon Valley challenge 2000: Year 4 report*. San Jose, CA: Joint Venture, Silicon Valley Network.

Laynea, A., Jules, V., Kutnick, P., & Layne, C. (2008). Academic achievement, pupil participation and integration of group work skills in secondary school classrooms in Trinidad and Barbados. *International Journal of Educational Development*, *28*(2), 176–194. doi:10.1016/j.ijedudev.2007.01.001

Pucher, R. K., Mense, A., Wahl, H., & Schmöllebeck, F. (2003). Intrinsic motivation of students in project based learning. *Transactions of the South African Institute of Electrical Engineers*, *94*(3), 6–9.

Railsback, J. (2002). *Project-based instruction: Creating excitement for learning*. Portland, OR: Northwest Regional Educational Laboratory.

Swan, C. W., Paterson, K. G., & Bielefeldt, A. R. (2009, October 18-21). Panel- measuring the impacts of project-based service learning in engineering education. In *Proceedings of the 39th ASEE/IEEE Frontiers in Education Session Conference*, San Antonio, TX.

Thomas, J. W. (1998). *Project-based learning: Overview*. Novato, CA: Buck Institute for Education.

Thomas, J. W. (2000). *A review of research on project-based learning*. San Rafael, CA: AutoDesk Foundation.

Van den Bergh, V., Mortelmans, D., Spooren, P., Van Petegem, P., Gijbels, D., & Vanthournout, G. (2006). New assessment modes within project-based education – the stakeholders. *Studies in Educational Evaluation*, *32*(4), 345–368. doi:10.1016/j.stueduc.2006.10.005

Wac-Włodarczyk, A., & Billewicz, P. (2008, November 4-6). Quality assurance procedures and methods of quality assessment across the European higher education area. In *Proceedings of the 8th Annual SEAAIR Conference*, Surabaya, Indonesia (pp. 1-11).

Wolfs, P. J., Howard, P., Vann, A., & Edwards, R. (1997). A response to the national engineering education review, BE2001: A project-based engineering degree. In *Proceedings of the Third Bi-annual Conference of the Australian Problem Based Learning: Research and Development in Problem Based Learning*, New South Wales, Australia (Vol. 4, pp. 674-681).

This work was previously published in International Journal of Quality Assurance in Engineering and Technology Education (IJQAETE), Volume 1, Issue 2, edited by Arun Patil, pp. 23-34, copyright 2011 by IGI Publishing (an imprint of IGI Global).

Section 3
Tools and Technologies

This section presents an extensive coverage of various tools and technologies available in the field of Industrial Engineering that practitioners and academicians alike can utilize to develop different techniques. These chapters enlighten readers about fundamental research on the many tools facilitating the burgeoning field of Industrial Engineering. It is through these rigorously researched chapters that the reader is provided with countless examples of the up-and-coming tools and technologies emerging from the field of Industrial Engineering. With 14 chapters, this section offers a broad treatment of some of the many tools and technologies within the Industrial Engineering field.

Chapter 28
Semantic Technologies in Motion:
From Factories Control to Customer Relationship Management

Ricardo Colomo-Palacios
Universidad Carlos III de Madrid, Spain

ABSTRACT

Semantic technologies are evolving to reach a mature state. Given the importance of Information Technologies in general and Internet in particular to organizations worldwide, these technologies can provide enhanced functionality to both existing and future solutions. This chapter provides an overview of the use of semantic technologies in several application domains, namely: construction & real estate, customer relationship management, e-government, e-learning, environmental sciences, health domain, human resource management, Information Technology, manufacturing, media and tourism, and cultural heritage. More than a hundred references are provided to support the idea that semantic technologies are applicable to almost all areas of interest.

INTRODUCTION

In recent years, the information technologies (IT) field has expanded beyond traditional stove-pipe organizational systems to infiltrate companies and homes alike (García Crespo et al., 2008). Today, the contribution of IT to economies is unquestioned (Stehr, 2007). IT has been considered fundamental for the development of productivity and knowledge-intensive products and services (Soto-Acosta et al., 2010). Moreover, the Internet-driven networked economy is evolving to the point where businesses are fully aware of the enormous business opportunities of online transactions (García-Crespo et al., 2011a) and such importance in recent years has turned their development into

DOI: 10.4018/978-1-4666-1945-6.ch028

a critical task for corporations (García-Crespo et al., 2009a). Given that organizations must continuously innovate in terms of product, process, market and business model to remain sustainable (O'Sullivan & Dooley, 2010), many organizations rely their innovation process and, in many cases, the whole of their business to IT systems. Advances in technology, emergence of new business practices, and shifting social and geopolitical circumstances have combined to create a "brave new IT world" for organizations (Goles, Hawk & Kaiser, 2008).

In this scenario, organizations use multiple IT/IS solutions to support their activities at all management levels (Trigo, Varajao & Barroso, 2009) and this is so because IT-enabled services makes possible to overcome geographical, temporal and organizational barriers to communication and knowledge transfer (Corso, Giacobbe & Martini, 2009). Moreover, Information technology organisations, who frequently lead the development of change based around ICT, are being asked to develop new products and services that add significant value for customers and to radically change their internal processes so that they are more cost effective (O'Sullivan & Dooley, 2010). One of the main strengths of the Web is that it allows any party of its global community to share information with any other party (Presutti & Gangemi, 2008). In an environment of globalization and competition, institutions have turned to knowledge as a strategic asset that drives sustainable economic advantage (Sharma et al., 2010) in which technology is the means in which enterprises collaborate all over the world. In this environment, the power of knowledge in our contemporary life has produced many new terms and concepts including: "knowledge society"; "knowledge economy", "knowledge management", and "knowledge culture" (Bakry & Alfantookh, 2010).

Since its humble beginnings, the Internet has gained vast importance in today's society, both in terms of consumer reach and the volume of fundamental information it contains for millions of users worldwide. Transforming from a host-to-host network connection in ARPANET (Abbate, 1999) on October 29, 1969, it has advanced to become what Tim Berners-Lee has termed the "Giant Global Graph" (Berners-Lee, 2007). As stated by Bieber et al. (1997), one of the original success factors of the Web is that it provides simple access to the information contained in it. Because of the power of IT, human kind has been turning its concentration on developing web-based services (Sudhahar et al., 2010). Moreover, according to Targowski (2009) Web technology is key solution for the provision of e-Service systems. When we talk about the development of the web and where we stand today, we stumble across two main buzzwords: Web 2.0 and Semantic Web (Lux & Dösinger, 2007).

The evolution of the Web from Web 1.0 to Web 2.0, and to what has recently been termed by some as Web 3.0, has caused a paradigm shift in a user's access to and control of websites. The Web 2.0 phenomenon made the Web social, initiating an explosion in the number of users of the Web, thus empowering them with a huge autonomy in adding content to web pages, labeling the content, creating folksonomies of tags, and finally, leading to millions of users constructing their own web pages (Breslin & Decker, 2007). Logically, the result of this movement was a significant increase in the number of web pages available. Surfing through endless links is no longer an efficient method for finding relevant information. This not only applies to the standard user, all owners of web pages, whether organizations or consumers, lose target audiences as users of the Web are unable to access their pages in an efficient time frame (García-Crespo et al., 2010a). Both the information that is stored on the Web and the number of its human users have been growing exponentially in recent years and, for many people, the Web has started to play a fundamental role as a means of providing and searching for information (Eiter et al., 2008). In this environment, both user computing and IT in organizations can benefit from the application of semantic technologies that enable the access and storage of information in an unified way. Not in vain, this technology promises to solve many challenging and cost-intensive problems of the current generation of Web applications (Nixon et al., 2008).

This chapter presents the implications of Semantic Technologies in several areas of interest, giving a sight of the main efforts devoted to the adaptation of semantic technologies to organizations and human activities in literature.

OUTLINE: SEMANTIC TECHNOLOGIES

Semantic Technologies have been pointed out as the future of Web (Benjamins et al., 2008) and a new way to support knowledge (Vossen et al., 2007; Fensel & Musen, 2001) in a wide range of domains (Lytras & García, 2008). Semantic technologies, based on ontologies (Fensel, 2002), provide a common framework that enables for data integration, sharing and reuse from multiple sources. Durguin and Sherif (2008) portrays the Semantic Web as the future web where computer software agents can carry out sophisticated tasks for users.

This approach facilitates the integration of data coming from a broader non-relational domain of data, which additionally might be distributed and outside enterprise boundaries and control (García, 2010a). Taking this into account, according to Alani et al., (2008), Semantic Web applications are beginning to be pragmatic. Technology Journalist Markoff (2006) begun to call this new web applications as Web 3.0 and this tendency was latter followed by others (e.g. Lassila & Hendler, 2007; Hendler, 2008; Wang, 2008; Hendler, 2009). Web 3.0 can bring a new breed of spectacular applications compared to Web 2.0 with the same magnitude that separates Web 2.0 from Web 1.0 (Cardoso, 2007).

Semantic Technologies have emerged as a new and highly promising context for knowledge and data engineering (Vossen et al., 2007). The term "Semantic Web" was coined by Berners-Lee, Hendler and Lassila (2001), to describe the evolution from a document-based web towards a new paradigm that includes data and information for computers to manipulate. The essential difference between the classic Web and the Semantic Web is that structured data is exposed in a structured way (Gruber, 2008).

The Semantic Web enables automated information access based on machine-processable semantics of data. Being machine-processable means that semantic search services can be made this information available for providing precise and exhaustive information retrieval (Guha, 2003). Ontologies provide information systems with a semantically rich knowledge base for the interpretation of unstructured content (Mikroyannidis & Theodoulidis, 2010).

According to Lytras and García (2008), in recent years, Semantic Web research has resulted in significant outcomes and the adoption of this technology from the market and the industry is becoming closer. Moreover, according to Ding (2010), Semantic Web is fastmoving in a multidisciplinary way. The Semantic Web provides an alternative solution to represent the comprehensive meaning of integrated information and promises to lead to efficient data management by establishing a common understanding by ontologies (Shadbolt et al., 2006). Ontologies (Fensel, 2002) are the technological cornerstones of the Semantic Web, because they provide structured vocabularies that describe a formal specification of a shared conceptualization. Ontologies were developed in the field of Artificial Intelligence to facilitate knowledge sharing and reuse (Fensel et al., 2001). An ontology can be defined as "a formal and explicit specification of a shared conceptualisation" (Studer et al., 1998). Ontologies provide a common vocabulary for a domain and define, with different levels of formality, the meaning of the terms and the relations between them. Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, functions, axioms and instances (Gruber, 1993). The theory which supports the use of ontologies is a formal theory within which not only definitions but also a supporting framework of axioms is included (Smith, 2003). Besides Semantic Web, ontology defines a set of representational primitives with which a domain of knowledge is modeled. Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, functions, axioms and instances (Gruber, 1993). Classes in the ontology are usually organized into taxonomies. Sometimes the definition of ontologies has been diluted, in the sense that taxonomies are considered to be full ontologies (Studer et al., 1998). Languages such as RDF and OWL have been developed; these languages allow for the description of web resources, and for the representation of knowledge that will enable applications to use resources more intelligently (Horrocks, 2008). These languages, and the tools developed to support them, have rapidly become de facto standards for ontology development and deployment; they are increasingly used, not only in research labs, but in large scale IT projects (Horrocks, 2008). The Semantic Web consists of several hierarchical layers, where the Ontology layer, in form of the OWL Web Ontology Language (recommended by the W3C), is currently the highest layer of sufficient maturity (Lukasiewicz & Straccia, 2008).

Taking full advantage of ontologies, the Semantic Web provides a complementary vision as a knowledge management environment (Warren, 2006) that, in many cases has expanded and replaced previous knowledge and information management archetypes (Davies, Lytras & Sheth, 2007). Thus, Semantic Web has emerged to be a new and highly promising context for knowledge and data engineering (Vossen, Lytras & Koudas, 2007). The goals of the Semantic Web initiative include the integration of data from different sources in a machine processable format in order to make them accessible to computer programs and facilitating the use of data in ways that have not been thought of when the data was entered or recorded (Battré, 2008). It is agreed that semantic enrichment of resources would lead to better search results (Scheir, Lindstaedt & Ghidini, 2008).

Having outlined the importance of Semantic Technologies, these technologies has recently attracted much attention, both from academia and industry, and is widely regarded as the next step in the evolution of the World Wide Web (Lukasiewicz & Straccia, 2008) and this recent boom in Semantic Web technologies has been occurring in the so-called "Web 3.0" technologies (Hendler & Golbeck, 2008).

SEMANTIC TECHNOLOGIES IN ORGANIZATIONS

In the past years our community has seen continuous, sustained growth in the deployment of Semantic Web inventiveness in large and small organizations of many types (Cardoso, 2007) and as a result of this, various industries have adopted SW technologies (Lytras & García, 2008). According to Breslin et al. (2010), industry has begun to watch developments with interest and a number of large companies have started to experiment with Semantic technologies to ascertain if these new technologies can be leveraged to add more value for their customers or internally within the company, while there are already several offers of vendors of Semantic solutions on the market. These authors stated that Semantic technologies are not only valuable on an open environment like the Web, but also in closed systems such as in industrial settings.

The aim of this section is to illustrate the use of semantic technologies in industries through a state of the art study in which for every area of application several works are outlined in which authors employ semantic technologies. In Table 1, these works are cited ordered by area of application and issue covered.

Table 1. Semantic technologies efforts ordered by area of application

Business Integration	Business integration tools	Arroyo et al. (2007)
Construction & Real Estate	Cadastrial domain application	Hess and de Vries (2006)
	Corporate memory management in building construction	El-Diraby and Zhang (2006)
	Real Estate Product Finder	Langegger and Wöß (2007)
	Communication in construction	Zhang and El-Diraby (2009)
	Domain Ontology for Processes in Infrastructure and Construction	El-Gohary, and El-Diraby (2010)
	Retail Real Estate Agencies Information System	García-Crespo et al (2010e)
Customer Relationship Management	Employee driven CRM ontolgy	Van Damme, Christiaens & Vandijck, (2007)
	Customer complaint management	Jarrar, (2008)
	Knowledge for Customer Relationship Management	Magro and Goy, (2008)
	Emotions analysis	García-Crespo et al (2009d)
	Customer social networks analysis	García-Crespo et al (2010f)
	State of the art	García-Crespo et al (2010g)
eGovernment	Data models and interfaces to eGovernment services:	Álvarez Sabucedo et al. (2009a)
	eGovernment interoperability	Peristeras, Tarabanis and Goudos (2009)
	Multi-agent based framework for e-government	Mellouli and Bouslama (2009)
	Operational procedures facilitation in public administration	Savvas & Bassiliades (2009)
	Provision of Services	Álvarez Sabucedo et al. (2009b)
	Semantic interoperability	Guijarro (2009)
	eGovernment agents	Álvarez Sabucedo et al. (2010)
	Information Processes	Klischewski and Ukena (2010)
eLearning	Dynamic e-Learning	Lytras, Pouloudi and Poulymenakou (2002).
	Knowledge Management Learning Activities	Sicilia et al. (2006)
	IMS Learning design	Naeve, Sicilia and Lytras (2008)
	Learning content usage analysis	Jovanovic et al. (2008)
	Learning in various domains	D'Avanzo, Kuflik and Lytras (2008)
	Ontology to Automate Learning Scenarios	Rius, Sicilia and García-Barriocanal (2008)
	Personalised learning services	Vargas-Vera & Lytras (2008)
	Developing semantic web-based educational systems	Bittencourt et al. (2009)
	Intelligent techniques in eLearning	Gladun et al. (2009)
	Interoperability of Learning Objects Copyright	García and Pariente (2009)
	Ontology-based training	Macris (2009)
	Recommendation methods to efficiently personalise resources	Kerkiri, Manitsaris and Mavridis (2009)
	Semantic web technology based framework for educational-offer selection	Kumar, Kumar and Jain (2009)
	Support Learning Content Development	Pahl & Holohan (2009)

continued on following page

Table 1. Continued

	Teaching oriented ontologies	Belesiotis and Alexandris, (2009)
	Feasibility Studies Preparation Training Process	Georgakellos and Macris (2010)
	Learning designs	Sicilia (2006)
	Agile process learning	Amescua et al. (2010)
Environmental science	Organic Food Industry	Brewster (2005)
	Fractures analysis	Zhong, Aydina & McGuinness (2009)
	Coral Reef Ecosystems Research	Myers, Atkinson and Johnstone (2010)
	Decision support system for Digital Earth	Zhang, Zhao & Li (2010a)
	Early Warning Systems	Casola et al. (2010)
	Search of geospatial features for disaster and emergency management	Zhang, Zhao & Li (2010b)
	Spatial data discovery:	Batcheller & Reitsma (2010)
	Urban Applications	Della Valle, Celino and Dell'Aglio (2010)
	Web services for rural areas	Pimenidis and Georgiadis (2010)
Health domain	Interoperability in heterogeneous medical information systems	Orgun & Vu (2006)
	Medical ontologies integration	Lee, Supekar and Geller (2006)
	Pharmaceutical product development and manufacturing	Venkatasubramanian et al. (2006)
	Semantically enriched Web services in the healthcare domain	Dogac et al. (2006)
	Knowledge-Centric Clinical Decision Support Systems	Hussain et al., (2007)
	Integrated access to biological resources	García-Sánchez et al. (2008)
	State of the art	Stenzhorn et al. (2008)
	Decision support system for pharmaceutical product development	Akkisetty, Reklaitis and Venkatasubramanian (2009)
	Knowledge management in biomedical libraries	Fuentes-Lorenzo, Morato and Gómez-Berbís (2009)
	Knowledge management in pharmaceutical industry	Pappa, Stergioulas and Telonis (2009)
	Ontology-based clinical Knowledge Management systems	Sicilia et al. (2009a)
	Semantic integration of data resources	Podgorelec, Grasic and Pavlic (2009)
	State of the art	Lytras, M.D., Sakkopoulos, E and Ordóñez-De Pablos (2009)
	State of the art	Schulz et al. (2009)
	Diagnosis advisor enabled by semantics	García-Crespo et al. (2010e)
	Differential diagnosis based on logical inference	García-Crespo et al. (2010d)
	Ontologies	Deogun & Spaulding (2010)
	Semantic Portals with health information	Suominen et al., (2009)
	Information Retreival of medical documents	Valencia-García et al. (2008)
	Clinical processes support	Serna Nocedal, Gerrikagoitia & Huerga (2010)
	Clinical trial – patient matching	Patel et al., (2010)
	Disease monitoring	Lagares-Lemos et al., (2011)
Human Resource Management	Knowledge management	Mika & Akkermans (2004)

continued on following page

Table 1. Continued

	Competence development	Dodero et al. (2007)
	Enterprise performance and competences monitoring	Jussupova-Mariethoz and Probst (2007)
	Task Assigment	Colucci et al. (2007)
	Assist work assignment	Macris, Papadimitriou and Vassilaco-poulos (2008)
	Competence Management	Draganidis, Chamopoulou & Mentzas (2008)
	Knowledge sharing and reuse	Lanzenberger (2008)
	Recruitment and personnel development assisted by semantics	Colomo-Palacios et al. (2008a)
	Assist mentoring processes	Colomo-Palacios et al. (2009)
	Competence Management	Hodik, Vokrinek &Becvar (2009)
	Team formation	Wi, Oh, Mun, Jung (2009)
	Competence evidence identification	Colomo-Palacios et al. (2010b)
	Competency-centric HRM	De Leenheer, Christiaens and Meers-man (2010)
	Knowledge classification	ZadJabbari, Wongthongtham & Hus-sain (2010)
	Knowledge development	Vargas-Vera & Lytras (2008)
	Technical competence assessment is software development teams	Colomo-Palacios et al. (2010a)
	Question answering	Vargas-Vera & Lytras (2010a); Vargas-Vera & Lytras (2010b)
	Mentoring	Soto-Acosta, Casado-Lumbreras & Cabezas-Isla (2010)
Information Technology	Change Management	Tripathi, Hinkelmann and Feldkamp (2008)
	Semantics-based social network for software projects	Gómez-Berbís et al. (2008)
	Software artifacts repository creation and development	Colomo-Palacios et al. (2008b)
	Software Reuse Support	Shiva and Shala (2008)
	Trazability among software artifacts	Zhang, Witte, Rilling, and Haarslev (2008)
	Use cases and patterns,	Dietrich, Jones and Wright (2008)
	Benchmarking metrics for software projects	García-Crespo et al. (2009c)
	Representation of tactile information for software development	Myrgioti, Chouvardas and Miliou (2009)
	Software engineering artifacts modeling	Sicilia et al. (2009b)
	Semantic-Web-Based Software Engineering	Zhao, Dong and Peng (2009)
	Collaborative Software Development	Happel, Maalej and Seedorf (2010)
	Quality Management	García et al. (2010b)
	Software analysis	Tappolet, Kiefer and Bernstein (2010)
	User Modelling	Sosnovsky & Dicheva (2010)
	Software Development teaming	Valencia-García et al. (2010)
	Modelling controlled flexibility in software processes	Martinho, Varajao & Domingos (2010)
	Formal methods in Software engineering	Krämer & Margaria (2009)
Manufacturing	Inter-enterprise collaboration in manufacturing	Lin, Hardin & Shahbaz (2004)

continued on following page

Table 1. Continued

	Collaborative engineering design	Zhang and Yin (2007)
	Inter-enterprise collaboration in manufacturing	Lin & Hardin (2007)
	Automotive supplier	Blomqvist & Öhgren (2008)
	Simulation component reuse	Bell et al. (2008).
	Industrial manufacturing processes	García-Crespo et al. (2010b)
	Product Lifecycle management	Matsokis and Kiritsis (2010)
Media	Newspapers	García et al. (2008)
	Marketing communication for media channels	Mulherna (2009)
	Scientific publishing	Shotton (2009)
	Financial headline news classification	Mellouli, Bouslama & Akande (2010)
	Movie scene characterization	Soleymani et al. (2009)
	Automatic music genre classification	Silla, Koerich & Kaestner (2099)
Research & Development	Managing of R& D Teams	Colomo-Palacios et al., (2010c)
Tourism & Cultural Heritage	Business integration in tourism using ontologies for mediation	Fodor and Werthner (2004)
	Semantically enriched travel Web services deployment	Doag et al. (2004)
	Ontology-based e-Tourism Planner	Jakkilinki et al. (2007)
	Semantic Web Service for Tourism Information over the Mobile Web	Lee (2007)
	Hotel search enhancement with Semantic Web Technologies	Niemann et al. (2008)
	Semantic-based information for tourism destinations	Kanellopoulos and Panagopoulos (2008)
	Use of ontologies in e-Tourism domain (matchmaking in Group Package Tours)	Kanellopoulos, (2008)
	Personalized recommendations for tourist attractions over the internet	Huang and Bian (2009)
	Provide recommendation in the context of the city of Taiwan	Lee, Chang and Wang (2009)
	Semantically enriched recommendation platform for tourists on route	García-Crespo et al. (2009a)
	Tourist service coordination	Chiu et al. (2009)
	Semantically enriched recommendation platform for tourists on route extended to Destination management organizations	García-Crespo et al. (2010c)
	Hotel recommendation	García-Crespo et al. (2011b)

CONCLUSION AND FUTURE WORK

Semantic technologies are beginning to be good shape. These technologies can be efficiently deployed for domains including Web Services, Enterprise Application Integration, Knowledge Management and E-Commerce, fulfilling existing gaps in current applications (Breslin et al., 2010). In this chapter, author has reviewed the use of Semantic Technologies in diverse industry and services domain. From the technological and research point of view it's clear that semantic technologies are mature, but their use in industrial settings is far from being extensive. Too few vendors bet for these technologies and include them into their products or solutions, while there is a legion of researchers around the world nurturing this phenomenon: producing thousands of papers, assisting to hundreds of conferences and even starting and editing a handful of good journals.

This chapter, as a part of the book devoted to CRM and Semantic Technologies, is aimed to prove that these technologies are applicable is practically every business domain.

REFERENCES

Abbate, J. (1999). *Inventing the Internet*. Cambridge, UK: MIT Press.

Akkisetty, P. K., Reklaitis, G. V., & Venkatasubramanian, V. (2009). Ontological informatics based decision support system for pharmaceutical product development: Milling as a case study. *Computer Aided Chemical Engineering*, *26*, 159–164. doi:10.1016/S1570-7946(09)70027-6

Alani, H., Hall, W., O'Hara, K., Shadbolt, N., Szomszor, M., & Chandler, P. (2008). Building a pragmatic Semantic Web. *IEEE Intelligent Systems*, *23*(3), 61–68. doi:10.1109/MIS.2008.42

Álvarez Sabucedo, L., Anido-Rifón, L., Corradini, F., Polzonetti, A., & Re, B. (2010). Knowledge-based platform for e-government agents: A Web-based solution using semantic technologies. *Expert Systems with Applications*, *37*(5), 3647–3656. doi:10.1016/j.eswa.2009.10.026

Álvarez Sabucedo, L., Anido-Rifón, L., Míguez-Pérez, R., & Santos-Gago, J. M. (2009a). Providing standard-oriented data models and interfaces to e-government services: A semantic-driven approach. *Computer Standards & Interfaces*, *31*(5), 1014–1027. doi:10.1016/j.csi.2008.09.042

Álvarez Sabucedo, L., Anido-Rifón, L., Míguez-Pérez, R., & Santos-Gago, J. M. (2009b). A holistic semantic framework for the provision of services in the domain of e-government. *International Journal of Software Engineering and Knowledge Engineering*, *19*(7), 961–993. doi:10.1142/S0218194009004490

Amescua, A., Bermón, L., García, J., & Sánchez-Segura, M. I. (2010). Knowledge repository to improve agile development processes learning. *IET Software*, *4*(6), 434–444. doi:10.1049/iet-sen.2010.0067

Arroyo, S., Sicilia, M. A., & Dodero, J. M. (2007). Choreography frameworks for business integration: Addressing heterogeneous semantics. *Computers in Industry*, *58*(6), 487–503. doi:10.1016/j.compind.2006.10.002

Bakry, S. H., & Alfantookh, A. (2010). Toward building the knowledge culture: Reviews and a KC-STOPE with Six Sigma view. *International Journal of Knowledge Society Research*, *1*(1), 47–65. doi:10.4018/jksr.2010010104

Batcheller, J. K., & Reitsma, F. (2010). Implementing feature level semantics for spatial data discovery: Supporting the reuse of legacy data using open source components. *Computers, Environment and Urban Systems*, *34*(4), 333–334. doi:10.1016/j.compenvurbsys.2010.02.008

Battré, D. (2008). Caching of intermediate results in DHT-based RDF stores. *International Journal of Metadata*. *Semantics and Ontologies*, *4*(3), 183–195.

Belesiotis, V. S., & Alexandris, N. (2009). A scenario for the development and use of teaching oriented ontologies. *International Journal of Metadata*. *Semantics and Ontologies*, *4*(3), 183–195. doi:10.1504/IJMSO.2009.027753

Bell, D., Mustafee, N., de Cesare, S., Lycett, M., Taylor, S. J. E., & Fishwick, P. A. (2008). Ontology engineering for simulation component reuse. *International Journal of Enterprise Information Systems*, *4*(4), 47–61. doi:10.4018/jeis.2008100104

Benjamins, V. R., Davies, J., Baeza-Yates, R., Mika, P., Zaragoza, H., & Greaves, M. (2008). Near-term prospects for semantic technologies. *IEEE Intelligent Systems*, *23*(1), 76–88. doi:10.1109/MIS.2008.10

Berners-Lee, T. (2007). *Giant global graph*. Retrieve September 9, 2010, from http://dig.csail.mit.edu/breadcrumbs/node/215

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43. doi:10.1038/scientificamerican0501-34

Bieber, M., Vitali, F., Ashman, H., Balasubramanian, V., & Oinas-Kukkonen, H. (1997). Fourth generation hypermedia: Some missing links for the World Wide Web. *International Journal of Human-Computer Studies*, *47*(1), 31–65. doi:10.1006/ijhc.1997.0130

Bittencourt, I. I., Costa, E., Silva, M., & Soares, E. (2009). A computational model for developing Semantic Web-based educational systems. *Knowledge-Based Systems*, *22*(4), 302–315. doi:10.1016/j.knosys.2009.02.012

Blomqvist, E., & Öhgren, A. (2008). Constructing an enterprise ontology for an automotive supplier. *Engineering Applications of Artificial Intelligence*, *21*(3), 386–439. doi:10.1016/j.engappai.2007.09.004

Breslin, J. G., & Decker, S. (2007). The future of social networks on the Internet: The need for semantics. *IEEE Internet Computing*, *11*(6), 86–90. doi:10.1109/MIC.2007.138

Breslin, J. G., O'Sullivan, D., Passant, A., & Vasiliu, L. (2010). Semantic Web computing in industry. *Computers in Industry*, *61*(8), 729–741.. doi:10.1016/j.compind.2010.05.002

Brewster, C. (2005). *AKTivefood: Semantic Web based knowledge conduits for the organic food industry*. In Workshop on Semantic Web Case Studies and Best Practices for eBusiness (SWCASE'05), Galway, Ireland.

Cardoso, J. (2007). The Semantic Web vision: Where are we? *IEEE Intelligent Systems*, *22*(5), 22–26. doi:10.1109/MIS.2007.4338499

Casola, V., D'Onofrio, L., Di Lorenzo, G., & Mazzocca, N. (2010). A service-based architecture for the interoperability of heterogeneous sensor data: A case study on early warning. In Konecny, M., Zlatanova, S., & Bandrova, T. L. (Eds.), *Geographic information and cartography for risk and crisis management, part 2* (pp. 249–263). doi:10.1007/978-3-642-03442-8_17

Chiu, D. K. W., Yueh, Y. T. F., Leung, H. F., & Hung, P. C. K. (2009). Towards ubiquitous tourist service coordination and process integration: A collaborative travel agent system architecture with Semantic Web services. *Information Systems Frontiers*, *11*(3), 241–256. doi:10.1007/s10796-008-9087-2

Colomo Palacios, R., García Crespo, A., Gómez Berbís, J. M., Casado-Lumbreras, C., & Soto-Acosta, P. (2010a). SemCASS: Technical competence assessment within software development teams enabled by semantics. *International Journal of Social and Humanistic Computing*, *1*(3), 232–245. doi:10.1504/IJSHC.2010.032685

Colomo-Palacios, R., García-Crespo, Á., Soto-Acosta, P., Ruano-Mayoral, M., & Jiménez-López, D. (2010c). A case analysis of semantic technologies for R&D intermediation information management. *International Journal of Information Management*, *30*(5), 465–469. doi:10.1016/j.ijinfomgt.2010.05.012

Colomo Palacios, R., Gómez Berbís, J. M., García Crespo, A., & Mencke, M. (2009). SeMatching: Using semantics to perform pair matching in mentoring processes. Conference Paper In Visioning and Engineering the Knowledge Society - A Web Science Perspective. *Lecture Notes in Artificial Intelligence, 5736*.

Colomo-Palacios, R., Gómez-Berbís, J. M., García-Crespo, A., & Puebla-Sánchez, I. (2008b). Social global repository: Using semantics and social web in software projects. *International Journal of Knowledge and Learning*, *4*(5), 452–464. doi:10.1504/IJKL.2008.022063

Colomo Palacios, R., Ruano Mayoral, M., Gómez Berbís, J. M., & García Crespo, A. (2008a). Semantic competence pull: A semantics-based architecture for filling competency gaps in organizations. In García, R. (Ed.), *Semantic Web for business: Cases and applications*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-066-0.ch015

Colomo Palacios, R., Ruano-Mayoral, M., Soto-Acosta, P., & García-Crespo, A. (2010b). The war for talent: Identifying competences in IT professionals through semantics. *International Journal of Sociotechnology and Knowledge Development*, *2*(3), 26–36. doi:10.4018/jskd.2010070103

Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F., & Ragone, A. (2007). Semantic-based skill management for automated task assignment and courseware composition. *Journal of Universal Computer Science*, *13*(9), 1184–1212.

Corso, M., Giacobbe, A., & Martini, A. (2009). Rethinking knowledge management: The role of ICT and the rise of the virtual workspace. *International Journal of Learning and Intellectual Capital*, *6*(3), 272–292. doi:10.1504/IJLIC.2009.025045

D'Avanzo, E., Kuflik, T., & Lytras, M. D. (2008). Building and using domain ontologies for learning in various domains: A Semantic Web-Based learning perspective. *International Journal of Knowledge and Learning*, *4*(4), 329–348. doi:10.1504/IJKL.2008.022054

Davies, J., Lytras, M. D., & Sheth, A. P. (2007). Semantic-Web-based knowledge management. *IEEE Internet Computing*, *11*(5), 14–16. doi:10.1109/MIC.2007.109

De Leenheer, P., Christiaens, S., & Meersman, R. (2010). Business semantics management: A case study for competency-centric HRM. *Computers in Industry*, *61*(8), 760–775..doi:10.1016/j.compind.2010.05.005

Della Valle, E., Celino, I., & Dell'Aglio, D. (2010). The experience of realizing a Semantic Web urban computing application. *Transactions in GIS*, *14*(2), 163–181. doi:10.1111/j.1467-9671.2010.01189.x

Deogun, J. S., & Spaulding, W. (2010). Conceptual development of mental health ontologies. In Ras, Z. W., & Tsay, L.-S. (Eds.), *Advances in intelligent Information Systems: Studies in computational intelligence, 265* (pp. 299–333). Berlin/Heidelberg, DE: Springer.

Dietrich, J., Jones, N., & Wright, J. (2008). Using social networking and semantic web technology in software engineering – Use cases, patterns, and a case study. *Journal of Systems and Software*, *81*(12), 2183–2193. doi:10.1016/j.jss.2008.03.060

Ding, Y. (2010). Semantic Web: Who is who in the field — A bibliometric analysis. *Journal of Information Science*, *36*(3), 335–356. doi:10.1177/0165551510365295

Dodero, J. M., Sánchez-Alonso, S., & Frosch-Wilke, D. (2007). generative instructional engineering of competence development programmes. *Journal of Universal Computer Science*, *13*(9), 1213–1233.

Dogac, A., Kabak, Y., Laleci, G., Sinir, S., Yildiz, A., & Kirbas, S. (2004). Semantically enriched Web services for the travel industry. *SIGMOD Record*, *33*(3), 21–27. doi:10.1145/1031570.1031575

Dogac, A., Laleci, G. B., Kirbas, S., Kabak, Y., Sinir, S., Yidiz, A., & Gurcan, Y. (2006). Artemis: Deploying semantically enriched Web services in the healthcare domain. *Information Systems*, *31*(4-5), 321–339. doi:10.1016/j.is.2005.02.006

Draganidis, F., Chamopoulou, P., & Mentzas, G. (2008). A Semantic Web architecture for integrating competence management and learning paths. *Journal of Knowledge Management*, *12*(6), 121–136. doi:10.1108/13673270810913667

Durguin, J. K., & Sherif, J. S. (2008). The Semantic Web: A catalyst for future e-business. *Kybernetes*, *37*(1), 49–65. doi:10.1108/03684920810850989

Eiter, T., Ianni, G., Lukasiewicz, T., Schindlauer, R., & Tompitsa, H. (2008). Combining answer set programming with description logics for the Semantic Web. *Artificial Intelligence*, *172*(12-13), 1495–1539. doi:10.1016/j.artint.2008.04.002

El-Diraby, T. E., & Zhang, J. (2006). A semantic framework to support corporate memory management in building construction. *Automation in Construction*, *15*(4), 504–521. doi:10.1016/j.autcon.2005.07.005

El-Gohary, N. M., & El-Diraby, T. E. (2010). Domain ontology for processes in infrastructure and construction. *Journal of Construction Engineering and Management*, *136*(7), 730–744. doi:10.1061/(ASCE)CO.1943-7862.0000178

Fensel, D. (2002). *Ontologies: A silver bullet for knowledge management and electronic commerce*. Berlin, Germany: Springer.

Fensel, D., & Munsen, M. A. (2001). The Semantic Web: A brain for humankind. *IEEE Intelligent Systems*, *16*(2), 24–25. doi:10.1109/MIS.2001.920595

Fensel, D., van Harmelen, F., Horrocks, I., McGuinness, D. L., & Patel-Schneider, P. F. (2001). OIL: An ontology infrastructure for the Semantic Web. *IEEE Intelligent Systems*, *16*(2), 38–45. doi:10.1109/5254.920598

Fodor, O., & Werthner, H. (2004). Harmonise: A step toward an interoperable e-tourism marketplace. *International Journal of Electronic Commerce*, *9*(2), 11–39.

Fuentes-Lorenzo, D., Morato, J., & Gómez-Berbís, J. M. (2009). Knowledge management in biomedical libraries: A Semantic Web approach. *Information Systems Frontiers*, *11*(4), 471–480. doi:10.1007/s10796-009-9159-y

García, R. (2010a). Using the rhizomer platform for semantic decision support systems development. *International Journal of Decision Support System Technology*, *2*(1), 60–80. doi:10.4018/jdsst.2010101605

García, R., Gil, R., Gimeno, J. M., Granollers, T., López, J. M., Oliva, M., & Pascual, A. (2010b). A semantic wiki for quality management in software development projects. *IET Software*, *4*(6), 386–395. doi:10.1049/iet-sen.2010.0044

García, R., & Pariente, T. (2010). Interoperability of learning objects copyright in the luisa semantic learning management system. *Information Systems Management*, *26*(3), 252–261. doi:10.1080/10580530903018037

García, R., Perdrix, F., Gil, R., & Oliva, M. (2008). The Semantic Web as a newspaper media convergence facilitator. *Journal of Web Semantics*, *6*(2), 151–161. doi:10.1016/j.websem.2008.01.002

García-Crespo, A., Chamizo, J., Rivera, I., Mencke, M., Colomo-Palacios, R., & Gómez-Berbís, J. M. (2009b). SPETA: Social pervasive e-tourism advisor. *Telematics and Informatics*, *26*(3), 306–315. doi:10.1016/j.tele.2008.11.008

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., Chamizo, J., & Mendoza-Cembranos, M. D. (2010e). S-SoDiA: A semantic enabled social diagnosis advisor. *International Journal of Society Systems Science*, *2*(3), 242–254. doi:10.1504/IJSSS.2010.033492

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., Chamizo, J., & Rivera, I. (2010c). Intelligent decision-support systems for e-tourism: Using SPETA II as a knowledge management platform for DMOs and e-tourism service providers. *International Journal of Decision Support System Technology*, *2*(1), 35–47. doi:10.4018/jdsst.2010101603

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., & García-Sánchez, F. (2010a). SOLAR: Social link advanced recommendation system. *Future Generation Computer Systems*, *26*(3), 374–380. doi:10.1016/j.future.2009.07.008

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., & Mencke, M. (2009c). BMR: Benchmarking metrics recommender for personnel issues in software development projects. International. *Journal of Computational Intelligence Systems*, *2*(3), 257–267.

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., & Paniagua-Martín, F. (2010e). SERREA: A semantic management system for retail real estate agencies. *International Journal of Information Systems in the Service Sector*, *2*(2), 33–47. doi:10.4018/jisss.2010040103

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., & Paniagua-Martín, F. (2010g). Customer relationship management in social and Semantic Web environments. *International Journal of Customer Relationship Marketing and Management*, *1*(2), 1–10. doi:10.4018/jcrmm.2010040101

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., & Ruano-Mayoral, M. (2009a). A project management methodology for commercial software reengineering. *International Journal of Project Organisation and Management*, *1*(3), 253–267. doi:10.1504/IJPOM.2009.027538

García-Crespo, A., Colomo-Palacios, R., Gómez-Berbís, J. M., & Ruiz-Mezcua, B. (2010f). SEMO: A framework for customer social networks analysis based on semantics. *Journal of Information Technology*, *25*(2), 178–188. doi:10.1057/jit.2010.1

García Crespo, A., Colomo Palacios, R., Gómez Berbís, J. M., & Tovar Caro, E. (2008). The IT crowd: Are we stereotypes? *IEEE IT Professional*, *10*(6), 24–28. doi:10.1109/MITP.2008.134

García-Crespo, A., Colomo-Palacios, R., Mencke, M., & Gómez-Berbís, J. M. (2009d). CUSENT: Social sentiment analysis using semantics for customer feedback. In Lytras, M. D., & Ordoñez de Pablos, P. (Eds.), *Social Web evolution: Integrating semantic applications and Web 2.0 technologies*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-272-5.ch007

García-Crespo, Á., Gómez-Berbís, J. M., Colomo-Palacios, R., & Alor-Henández, G. (2011a). SecurOntology: A Semantic Web access control framework. *Computer Standards & Interfaces*, *33*(1), 42–49. doi:10.1016/j.csi.2009.10.003

García-Crespo, Á., López-Cuadrado, J. L., Colomo-Palacios, R., González-Carrasco, I., & Ruiz-Mezcua, B. (2011b). Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain. *Expert Systems with Applications*, *38*(10), 13310–13319. doi:10.1016/j.eswa.2011.04.152

García-Crespo, A., Rodríguez, A., Mencke, M., Gómez-Berbís, J. M., & Colomo-Palacios, R. (2010d). ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. *Expert Systems with Applications*, *37*(3), 2621–2628. doi:10.1016/j.eswa.2009.08.016

García-Crespo, A., Ruiz-Mezcua, B., López-Cuadrado, J. L., & Gómez-Berbis, J. M. (2010b). Conceptual model for semantic representation of industrial manufacturing processes. *Computers in Industry*, *61*(7), 595–612. doi:10.1016/j.compind.2010.01.004

García-Sánchez, F., Fernández-Breis, J. T., Valencia-García, R., Gómez, J. M., & Martínez-Béjar, R. (2008). Combining Semantic Web technologies with multi-agent systems for integrated access to biological resources. *Journal of Biomedical Informatics*, *41*(5), 848–859. doi:10.1016/j.jbi.2008.05.007

Georgakellos, D. A., & Macris, A. M. (2010). Application of the semantic learning approach in the feasibility studies preparation training process. *Information Systems Management*, *26*(3), 231–240. doi:10.1080/10580530903017708

Gladun, A., Rogushina, J., García-Sánchez, F., Martínez-Béjar, R., & Fernández-Breis, J. T. (2009). An application of intelligent techniques and Semantic Web technologies in e-learning environments. *Expert Systems with Applications*, *36*(2), 1922–1931. doi:10.1016/j.eswa.2007.12.019

Goles, T., Hawk, S., & Kaiser, K. M. (2008). Information Technology workforce skills: The software and IT services provider perspective. *Information Systems Frontiers*, *10*(2), 179–194. doi:10.1007/s10796-008-9072-9

Gómez Berbís, J. M., Colomo Palacios, R., García Crespo, A., & Ruiz Mezcua, B. (2008). ProLink: A semantics-based social network for software projects. *International Journal of Information Technology and Management*, *7*(4), 392–405. doi:10.1504/IJITM.2008.018656

Gruber, T. (2008). Collective knowledge systems: Where the Social Web meets the Semantic Web. *Web Semantics: Science*. *Services and Agents on the World Wide Web*, *6*(1), 4–13. doi:10.1016/j.websem.2007.11.011

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220. doi:10.1006/knac.1993.1008

Guha, R., McCool, R., & Miller, E. (2003). *Semantic search*. In WWW 2003: 12th Web Conference. Budapest, Hungary.

Guijarro, L. (2009). Semantic interoperability in e-government initiatives. *Computer Standards & Interfaces*, *31*(1), 174–180. doi:10.1016/j.csi.2007.11.011

Happel, H. G., Maalej, W., & Seedorf, S. (2010). Applications of ontologies in collaborative software development. *Collaborative Software Engineering, 1*, 109-129. Berlin/Heidelberg, Germany: Springer.

Hendler, J. (2008). Web 3.0: Chicken farms on the Semantic Web. *Computer*, *41*(1), 106–108. doi:10.1109/MC.2008.34

Hendler, J. (2009). Web 3.0 emerging. *Computer*, *42*(1), 111–113. doi:10.1109/MC.2009.30

Hendler, J., & Golbeck, J. (2008). Metcalfe's law, Web 2.0, and the Semantic Web. *Web Semantics: Science. Services and Agents on the World Wide Web*, *6*(1), 14–20. doi:10.1016/j.websem.2007.11.008

Hess, C., & de Vries, M. (2006). From models to data: A prototype query translator for the cadastral domain. *Computers, Environment and Urban Systems*, *30*(5), 529–542. doi:10.1016/j.compenvurbsys.2005.08.008

Hodik, J., Vokrinek, J., & Becvar, P. (2009). Support for virtual organisation creation – Partners' profiles and competency management. *International Journal of Agent-Oriented Software Engineering*, *3*(2/3), 230–251. doi:10.1504/IJAOSE.2009.023638

Horrocks, I. (2008). Ontologies and the Semantic Web. *Communications of the ACM*, *51*(12), 58–67. doi:10.1145/1409360.1409377

Huang, C. Y., & Lo, J. H. (2006). Optimal resource allocation for cost and reliability of modular software systems in the testing phase. *Journal of Systems and Software*, *79*(5), 653–664. doi:10.1016/j.jss.2005.06.039

Huang, Y., & Bian, L. (2009). A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet. *Expert Systems with Applications*, *36*(1), 933–943. doi:10.1016/j.eswa.2007.10.019

Hussain, S., Raza Abidi, S., & Raza Abidi, S. S. (2007). Semantic Web framework for knowledge-centric clinical decision support systems. In *Artificial Intelligence in Medicine* [). Berlin/Heidelberg, DE: Springer.]. *Lecture Notes in Computer Science*, *4594*, 451–455. doi:10.1007/978-3-540-73599-1_60

Jakkilinki, R., Georgievski, M., & Sharda, N. (2007). Connecting destinations with an ontology-based e-tourism planner. In Sigala, M., Mich, L., & Murphy, J. (Eds.), *Information and communication technologies in tourism 2007* (pp. 21–32). Berlin, Germany: Springer Wien. doi:10.1007/978-3-211-69566-1_3

Jarrar, M. (2008). Towards effectiveness and transparency in e-business transactions, an ontology for customer complaint management. In García, R. (Ed.), *Semantic Web for business: Cases and applications*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-066-0.ch007

Jovanovic, J., Gasevic, D., Brooks, C., Devedzic, V., Hatala, M., Eap, T., & Richards, G. (2008). LOCO-Analyst: Semantic Web technologies in learning content usage analysis. *International Journal of Continuing Engineering Education and Lifelong Learning*, *18*(1), 54–76. doi:10.1504/IJCEELL.2008.016076

Jussupova-Mariethoz, Y., & Probst, A. R. (2007). Business concepts ontology for an enterprise performance and competences monitoring. *Computers in Industry*, *58*(2), 118–129. doi:10.1016/j.compind.2006.09.008

Kanellopoulos, D. N. (2008). An ontology-based system for intelligent matching of travellers' needs for group package tours. *International Journal of Digital Culture and Electronic Tourism*, *1*(1), 76–99. doi:10.1504/IJDCET.2008.020136

Kanellopoulos, D. N., & Panagopoulos, A. (2008). Exploiting tourism destinations' knowledge in an RDF-based P2P network. *Journal of Network and Computer Applications*, *31*(1), 179–200. doi:10.1016/j.jnca.2006.03.003

Kerkiri, T., Manitsaris, A., & Mavridis, I. (2009). How e-learning systems may benefit from ontologies and recommendation methods to efficiently personalise resources. *International Journal of Knowledge and Learning*, *5*(3/4), 347–370. doi:10.1504/IJKL.2009.031229

Klischewski, R., & Ukena, S. (2010). E-government goes Semantic Web: How administrations can transform their information processes. *Semantic Technologies for E-Government,* 99-125. Berlin/Heidelberg, Germany: Springer.

Krämer, B. J., & Margaria, T. (2009). A hindsight on formal methods and prospects of semantic computing in software engineering. *International Journal of Semantic Computing*, *3*(1), 5–30. doi:10.1142/S1793351X09000641

Kumar, S., Kumar, K., & Jain, S. (2009). A Semantic Web technology based framework for educational-offer selection in higher education. *International Journal of Metadata*. *Semantics and Ontologies*, *4*(3), 165–182. doi:10.1504/IJMSO.2009.027752

Lagares-Lemos, Á., Lagares-Lemos, M., Colomo-Palacios, R., García-Crespo, Á., & Gómez-Berbís, J. M. (2011). DISMON: Using social Web and semantic technologies to monitor diseases in limited environments. *Journal of Information Technology Research*, *4*(1), 48–59. doi:10.4018/jitr.2011010104

Langegger, A., & Wöß, W. (2007). Product finding on the Semantic Web: A search agent supporting products with limited availability. *International Journal of Web Information Systems*, *3*(1/2), 61–88. doi:10.1108/17440080710829225

Lanzenberger, M., Sampson, J., Rester, M., Naudet, Y., & Latour, T. (2008). Visual ontology alignment for knowledge sharing and reuse. *Journal of Knowledge Management*, *12*(6), 102–120. doi:10.1108/13673270810913658

Lassila, O., & Hendler, J. (2007). Embracing Web 3.0. *IEEE Internet Computing*, *11*(3), 90–93. doi:10.1109/MIC.2007.52

Lee, C. S., Chang, Y. C., & Wang, M. H. (2009). Ontological recommendation multi-agent for Taiwan City travel. *Expert Systems with Applications*, *36*(1), 6740–6753. doi:10.1016/j.eswa.2008.08.016

Lee, Y., Supekar, K., & Geller, J. (2006). Ontology integration: experience with medical terminologies. *Computers in Biology and Medicine*, *36*(7-8), 893–919. doi:10.1016/j.compbiomed.2005.04.013

Lee, Y. W. (2007). A Semantic Web service for tourism information over the mobile web. *Journal of the Korean Geographical Society*, *42*(5), 788–805.

Lin, H. K., & Harding, J. A. (2007). A manufacturing system engineering ontology model on the Semantic Web for inter-enterprise collaboration. *Computers in Industry*, *58*(5), 428–437. doi:10.1016/j.compind.2006.09.015

Lin, H. K., Harding, J. A., & Shahbaz, M. (2004). Manufacturing system engineering ontology for semantic interoperability across extended project teams. *International Journal of Production Research*, *42*(24), 5099–5118. doi:10.1080/00207540412331281999

Lukasiewicz, T., & Straccia, U. (2008). Managing uncertainty and vagueness in description logics for the Semantic Web. *Web Semantics: Science. Services and Agents on the World Wide Web*, *6*(4), 291–308. doi:10.1016/j.websem.2008.04.001

Lux, M., & Dösinger, G. (2007). From folksonomies to ontologies: Employing wisdom of the crowds to serve learning purposes. *International Journal of Knowledge and Learning*, *3*(4/5), 515–528. doi:10.1504/IJKL.2007.016709

Lytras, M. D., Castillo-Merino, D., & Serradell-Lopez, E. (2010). New human resources practices, technology and their impact on SMEs' efficiency. *Information Systems Management*, *27*(3), 267–273. doi:10.1080/10580530.2010.493843

Lytras, M. D., & García, R. (2008). Semantic Web applications: A framework for industry and business exploitation - What is needed for the adoption of the Semantic Web from the market and industry. *International Journal of Knowledge and Learning*, *4*(1), 93–108. doi:10.1504/IJKL.2008.019739

Lytras, M. D., Pouloudi, A., & Poulymenakou, A. (2002). Dynamic e-learning setting through advanced semantics: The value justification of a knowledge management oriented metadata schema. *International Journal on E-Learning*, *1*(4), 49–61.

Lytras, M. D., Sakkopoulos, E., & Ordóñez-De Pablos, P. (2009). Semantic Web and knowledge management for the health domain: State of the art and challenges for the Seventh Framework Programme (FP7) of the European Union (2007-2013). *International Journal of Technology Management*, *47*(1-3), 239–249. doi:10.1504/IJTM.2009.024124

Macris, A., Papadimitriou, E., & Vassilacopoulos, G. (2008). An ontology-based competency model for workflow activity assignment policies. *Journal of Knowledge Management*, *12*(6), 72–88. doi:10.1108/13673270810913630

Macris, A. M. (2009). Enhancing enterprise resource planning users' understanding through ontology-based training. *International Journal of Knowledge and Learning*, *5*(5/6), 404–422. doi:10.1504/IJKL.2009.031506

Magro, D., & Goy, A. 2008. the business knowledge for customer relationship management: An ontological perspective. In *Proceedings of 1st International Workshop on Ontology-supported Business Intelligence* (OBI 2008). New York, NY: ACM.

Markoff, J. (2006, November 12). Entrepreneurs see a Web guided by common sense. *The New York Times*.

Martinho, R., Varajao, J., & Domingos, D. (2010). Using the semantic web to define a language for modelling controlled flexibility in software processes. *IET Software*, *4*(6), 396–406. doi:10.1049/iet-sen.2010.0045

Matsokis, A., & Kiritsis, D. (2010). An ontology-based approach for product lifecycle management. *Computers in Industry*, *61*(8), 787–797. doi:10.1016/j.compind.2010.05.007

Mellouli, S., & Bouslama, F. (2010). Multi-agent based framework for e-government. *Electronic Government, an International Journal, 6*(2), 177-192.

Mellouli, S., Bouslama, F., & Akande, A. (2010). An ontology for representing financial headline news. *Web Semantics: Science. Services and Agents on the World Wide Web*, *8*(2-3), 203–208. doi:10.1016/j.websem.2010.02.001

Mika, P., & Akkermans, H. (2004). Towards a new synthesis of ontology technology and knowledge management. *The Knowledge Engineering Review*, *19*(4), 317–345. doi:10.1017/S0269888905000305

Mikroyannidis, A., & Theodoulidis, B. (2010). Ontology management and evolution for business intelligence. *International Journal of Information Management*, *30*(6), 559–566. doi:10.1016/j.ijinfomgt.2009.10.002

Mulherna, F. (2009). Integrated marketing communications: From media channels to digital connectivity. *Journal of Marketing Communications*, *15*(2-3), 85–101. doi:10.1080/13527260902757506

Myers, T. S., Atkinson, I., & Johnstone, R. (2010). Supporting coral reef ecosystems research through modelling a reusable ontology framework. *Applied Artificial Intelligence*, *24*(1-2), 77–101. doi:10.1080/08839510903448676

Myrgioti, E. V., Chouvardas, V. G., & Miliou, A. N. (2009). Ontological representation of tactile information for software development. *Applied Ontology*, *4*(2), 139–167.

Naeve, A., Sicilia, M. A., & Lytras, M. D. (2008). Learning processes and processing learning: From organizational needs to learning designs. *Journal of Knowledge Management*, *12*(6), 5–14. doi:10.1108/13673270810913586

Niemann, M., Mochol, M., & Tolksdorf, R. (2008). Enhancing hotel search with Semantic Web technologies. *Journal of Theoretical and Applied Electronic Commerce Research*, *3*(2), 82–96.

Nixon, L. J. B., Simperl, E., Krummenacher, R., & Martin-Recuerda, F. (2008). Tuplespace-based computing for the Semantic Web: A survey of the state-of-the-art. *The Knowledge Engineering Review*, *23*(2), 181–212. doi:10.1017/S0269888907001221

O'Sullivan, D., & Dooley, L. (2010). Collaborative innovation for the management of Information Technology resources. *International Journal of Human Capital and Information Technology Professionals*, *1*(1), 16–30. doi:10.4018/jhcitp.2010091102

Orgun, B., & Vu, J. (2006). HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. *Computers in Biology and Medicine*, *36*(7-8), 817–836. doi:10.1016/j.compbiomed.2005.04.010

Pahl, C., & Holohan, E. (2009). Applications of Semantic Web technology to support learning content development. *Interdisciplinary Journal of E-Learning and Learning Objects*, *5*, 1–26.

Pappa, D. D., Stergioulas, L. K., & Telonis, P. (2009). The role of knowledge management in the pharmaceutical enterprise. *International Journal of Technology Management*, *47*(1-3), 127–144. doi:10.1504/IJTM.2009.024118

Patel, C., Gomadam, K., Khan, S., & Garg, V. (2010). TrialX: Using semantic technologies to match patients to relevant clinical trials based on their personal health records. *Web Semantics: Science*. *Services and Agents on the World Wide Web*, *8*(4), 342–347. doi:10.1016/j.websem.2010.08.004

Peristeras, V., Tarabanis, K., & Goudos, S. K. (2009). Model-driven e-government interoperability: A review of the state of the art. *Computer Standards & Interfaces*, *31*(4), 613–628. doi:10.1016/j.csi.2008.09.034

Pimenidis, E., & Georgiadis, C. K. (2010). Web services for rural areas—Security challenges in development and use. *Computers and Electronics in Agriculture*, *70*(2), 348–354. doi:10.1016/j.compag.2009.07.016

Podgorelec, V., Grasic, B., & Pavlic, L. (2009). Medical diagnostic process optimization through the semantic integration of data resources. *Computer Methods and Programs in Biomedicine*, *95*(2Suppl. 1), S55–S67. doi:10.1016/j.cmpb.2009.02.015

Presutti, V., & Gangemi, A. (2008). Identity of resources and entities on the Web. *International Journal on Semantic Web and Information Systems*, *4*(2), 49–72. doi:10.4018/jswis.2008040103

Rius, A., Sicilia, M. A., & García-Barriocanal, E. (2008). An ontology to automate learning scenarios? An approach to its knowledge domain. *Interdisciplinary Journal of E-Learning and Learning Objects*, *4*, 151–165.

Savvas, I., & Bassiliades, N. (2009). A process-oriented ontology-based knowledge management system for facilitating operational procedures in public administration. *Expert Systems with Applications*, *36*(1), 4467–4478. doi:10.1016/j.eswa.2008.05.022

Scheir, P., Lindstaedt, S. N., & Ghidini, C. (2008). A network model approach to retrieval in the Semantic Web. *International Journal on Semantic Web and Information Systems*, *4*(4), 56–84. doi:10.4018/jswis.2008100103

Serna Nocedal, A., Gerrikagoitia, J. K., & Huerga, I. (2010). Supporting clinical processes with Semantic Web technologies: A case in breast cancer treatment. *International Journal of Metadata*. *Semantics and Ontologies*, *5*(4), 309–320. doi:10.1504/IJMSO.2010.035553

Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The Semantic Web revisited. *IEEE Intelligent Systems*, *21*(3), 96–101. doi:10.1109/MIS.2006.62

Sharma, R. S., Ng, E. W. J., Dharmawirya, M., & Samuel, E. M. (2010). A policy framework for developing knowledge societies. *International Journal of Knowledge Society Research*, *1*(1), 23–46. doi:10.4018/jksr.2010010103

Shiva, S. G., & Shala, L. A. (2008). Using Semantic wikis to support software reuse. *Journal of Software*, *3*(4), 1–8. doi:10.4304/jsw.3.4.1-8

Shotton, D. (2009). Semantic publishing: The coming revolution in scientific journal publishing. *Learned Publishing*, *22*(2), 85–94. doi:10.1087/2009202

Sicilia, J. J., Sicilia, M. A., Sánchez-Alonso, S., García-Barriocanal, E., & Pontikaki, M. (2009a). Knowledge representation issues in ontology-based clinical knowledge management systems. *International Journal of Technology Management*, *47*(1-3), 191–206. doi:10.1504/IJTM.2009.024122

Sicilia, M. A. (2006). Semantic learning designs: Recording assumptions and guidelines. *British Journal of Educational Technology*, *37*(3), 331–350. doi:10.1111/j.1467-8535.2006.00609.x

Sicilia, M. A., García-Barriocanal, E., Sánchez-Alonso, S., & Rodríguez-García, D. (2009b). Ontologies of engineering knowledge: general structure and the case of Software Engineering. *The Knowledge Engineering Review*, *24*(3), 309–326. doi:10.1017/S0269888909990087

Sicilia, M. A., Lytras, M., Rodríguez, E., & García-Barriocanal, E. (2006). Integrating descriptions of knowledge management learning activities into large ontological structures: A case study. *Data & Knowledge Engineering*, *57*(2), 111–121. doi:10.1016/j.datak.2005.04.001

Silla, C. N., Koerich, A. L., & Kaestner, C. A. A. (2009). A feature selection approach for automatic music genre classification. *International Journal of Semantic Computing*, *3*(2), 183–208. doi:10.1142/S1793351X09000719

Smith, B. (2003). Ontology. An introduction. In Floridi, L. (Ed.), *Blackwell guide to the philosophy of computing and information* (pp. 155–166). Oxford, UK: Blackwell.

Soleymani, M., Chanel, G., Kierkels, J. J. M., & Pun, T. (2009). Affective characterization of movie scenes based on content analysis and physiological changes. *International Journal of Semantic Computing*, *3*(2), 235–254. doi:10.1142/S1793351X09000744

Sosnovsky, S., & Dicheva, D. (2010). Ontological technologies for user modeling. *International Journal of Metadata. Semantics and Ontologies*, *5*(1), 32–71. doi:10.1504/IJMSO.2010.032649

Soto-Acosta, P., Casado-Lumbreras, C., & Cabezas-Isla, F. (2010). Shaping human capital in software development teams: the case of mentoring enabled by semantics. *IET Software*, *4*(6), 445–452. doi:10.1049/iet-sen.2010.0087

Soto-Acosta, P., Martínez-Conesa, I., & Colomo-Palacios, R. (2010). An empirical analysis of the relationship between IT training sources and IT value. *Information Systems Management*, *27*(3), 274–283. doi:10.1080/10580530.2010.493847

Stehr, N. (2007). Societal transformations, globalisation and the knowledge society. *International Journal of Knowledge and Learning*, *3*(2/3), 139–153. doi:10.1504/IJKL.2007.015548

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1-2), 161–197. doi:10.1016/S0169-023X(97)00056-6

Sudhahar, C., Suresh Premil Kumar, R., Senthil, V., & Devadasan, S. R. (2010). WQC: A Web-based technique for solving organisational problems through global intellectual capital sharing. *International Journal of Learning and Intellectual Capital*, *7*(1), 75–97. doi:10.1504/IJLIC.2010.029535

Suominen, O., Hyvönen, E., Viljanen, K., & Hukka, E. (2009). HealthFinland-A national semantic publishing network and portal for health information. *Journal of Web Semantics*, *7*(4), 271–376. doi:10.1016/j.websem.2009.09.003

Tappolet, J., Kiefer, C., & Bernstein, A. (2010). Semantic Web enabled software analysis. *Web Semantics: Science. Services and Agents on the World Wide Web*, *8*(2-3), 225–240. doi:10.1016/j.websem.2010.04.009

Targowski, A. (2009). The architecture of service systems as the framework for the definition of service science scope. *International Journal of Information Systems in the Service Sector*, *1*(1), 54–77. doi:10.4018/jisss.2009010104

Trigo, A., Varajao, J., & Barroso, J. (2009). A practitioner's roadmap to learning the available tools for Information System Function management. *International Journal of Teaching and Case Studies*, *2*(1), 29–40. doi:10.1504/IJTCS.2009.026297

Tripathi, U. K., Hinkelmann, K., & Feldkamp, D. (2008). Life cycle for change management in business processes using semantic technologies. *Journal of Computers*, *3*(1), 24–31. doi:10.4304/jcp.3.1.24-31

Valencia-García, R., Fernández-Breis, J. T., Ruiz-Martínez, J. M., García-Sánchez, F., & Martínez-Béjar, R. (2008). A knowledge acquisition methodology to ontology construction for information retrieval from medical documents. *Expert Systems: The Knowledge Engineering Journal*, *25*(3), 314–334.

Valencia-García, R., García-Sánchez, F., Castellanos-Nieves, D., Fernández-Breis, J. T., & Toval, A. (2010). Exploitation of social semantic technology for software development team configuration. *IET Software*, *4*(6), 373–385. doi:10.1049/iet-sen.2010.0043

Van Damme, C., Christiaens, S., & Vandijck, E. (2007). Building an employee-driven CRM ontology. In *Proceedings of the IADIS Multi Conference on Computer Science and Information Systems* (pp.330-334). Lisbon, Portugal: E-society2007.

Vargas-Vera, M., & Lytras, M. D. (2008). Exploiting Semantic Web and ontologies for personalised learning services: Towards Semantic Web-enabled learning portals for real learning experiences. *International Journal of Knowledge and Learning*, *4*(1), 1–17. doi:10.1504/IJKL.2008.019734

Vargas-Vera, M., & Lytras, M. D. (2010a). AQUA: A Closed-Domain Question Answering System. *Information Systems Management*, *27*(3), 217–225. doi:10.1080/10580530.2010.493825

Vargas-Vera, M., & Lytras, M. D. (2010b). AQUA: Hybrid architecture for question answering services. *IET Software*, *4*(6), 418–433. doi:10.1049/iet-sen.2010.0053

Venkatasubramanian, V., Zhao, C., Joglekar, G., Jair, A., Hailemariam, L., & Suresh, P. (2006). Ontological informatics infrastructure for pharmaceutical product development and manufacturing. *Computers & Chemical Engineering*, *30*(10-12), 1482–1496. doi:10.1016/j.compchemeng.2006.05.036

Vossen, G., Lytras, M., & Koudas, N. (2007). Editorial: Revisiting the (Machine) Semantic Web: The missing layers for the human Semantic Web. *IEEE Transactions on Knowledge and Data Engineering*, *19*(2), 145–148. doi:10.1109/TKDE.2007.30

Wang, F. Y. (2008). Beyond X 2.0: Where should we go? *IEEE Intelligent Systems*, *24*(3), 2–4. doi:10.1109/MIS.2009.48

Warren, P. (2006). Knowledge management and the Semantic Web: From scenario to technology. *IEEE Intelligent Systems*, *21*(1), 53–59. doi:10.1109/MIS.2006.12

Zad Jabbari, B., Wongthongtham, P., & Hussain, F. K. (2010). Ontology based approach in knowledge sharing measurement. *Journal of Universal Computer Science*, *16*(6), 956–982.

Zhang, C., Zhao, T., & Li, W. (2010a). The framework of a geospatial Semantic Web-based spatial decision support system for Digital Earth. *International Journal of Digital Earth*, *3*(2), 111–134. doi:10.1080/17538940903373803

Zhang, C., Zhao, T., & Li, W. (2010b). Automatic search of geospatial features for disaster and emergency management. *International Journal of Applied Earth Observation and Geoinformation*, *12*(6), 409–418. doi:10.1016/j.jag.2010.05.004

Zhang, J., & El-Diraby, T. E. (2009). SSWP: A Social Semantic Web Portal for communication in construction. *Journal of Computers*, *4*(4), 330–337. doi:10.4304/jcp.4.4.330-337

Zhang, W. Y., & Yin, J. W. (2007). Exploring Semantic Web technologies for ontology-based modeling in collaborative engineering design. *International Journal of Advanced Manufacturing Technology*, *36*(9-10), 833–843. doi:10.1007/s00170-006-0896-5

Zhang, Y., Witte, R., Rilling, J., & Haarslev, V. (2008). Ontological approach for the semantic recovery of traceability links between software artefacts. *IET Software*, *2*(3), 185–203. doi:10.1049/iet-sen:20070062

Zhao, Y., Dong, J., & Peng, T. (2009). Ontology classification for Semantic-Web-Based software engineering. *IEEE Transactions on Services Computing*, *2*(4), 303–317. doi:10.1109/TSC.2009.20

Zhong, J., Aydina, A., & McGuinness, D. L. (2009). Ontology of fractures. *Journal of Structural Geology*, *31*(3), 251–259.

ADDITIONAL READING

Blumauer, A., & Pellegrini, T. (2009). *Social Semantic Web*. Springer. doi:10.1007/978-3-540-72216-8

Breslin, J. G., Passant, A., & Decker, S. (2010). *The Social Semantic Web*. Springer.

Cardoso, J., Hepp, M., & Lytras, M. D. (2008). *The semantic web: real-world application from industry*. Springer.

Cardoso, J., & Lytras, M. D. (2009). *Semantic Web Engineering in the Knowledge Society*. IGI-Global.

García, R. (2008). *Semantic Web for Business: Cases and Applications*. IGI-Global. doi:10.4018/978-1-60566-066-0

García-Castro, R. (2010). *Benchmarking Semantic Web Technology*. AKA Verlag Heidelberg and IOS Press.

Hitzler, P., Krötzsch, M., & Rudolph, S. (2009). *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC.

Kalfoglou, Y. (2010). *Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications*. IGI-Global.

Lytras, M. D., & Ordóñez-De Pablos, P. (2009). *Social Web Evolution: Integrating Semantic Applications and Web 2.0 Technologies*. IGI-Global. doi:10.4018/978-1-60566-272-5

Lytras, M. D., & Sheth, A. (2010). *Progressive Concepts for Semantic Web Evolution: Applications and Development*. IGI-Global. doi:10.4018/978-1-60566-992-2

Mika, P. (2009). *Social Networks and the Semantic Web*. Springer.

Murugesan, S. (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Application*. IGI-Global.

Naeve, A., & Lytras, M. D. (2006). *Intelligent Learning Infrastructure for Knowledge Intensive Organizations: A Semantic Web Perspective*. Idea Group.

Paquette, G. (2010). *Visual Knowledge Modeling for Semantic Web Technologies: Models and Ontologies*. IGI-Global. doi:10.4018/978-1-61520-839-5

Sheth, A., & Lytras, M. D. (2007). *Semantic Web-Based Information Systems: State-of-the-Art Applications*. Idea Group.

Tao, D., Xu, D., & Li, H. (2009). *Semantic Mining Technologies for Multimedia Databases*. IGI-Global. doi:10.4018/978-1-60566-188-9

Wi, H., Oh, S., Mun, J., & Jung, M. (2009). A team formation model based on knowledge and collaboration. *Expert Systems with Applications*, *36*(5), 9121–9134. doi:10.1016/j.eswa.2008.12.031

Zhang, Y. J. (2007). *Semantic-Based Visual Information Retrieval*. IGI-Global.

Zilli, A., Damiani, E., Ceravolo, P., Corallo, A., & Elia, G. (2009). *Semantic Knowledge Management: An Ontology-Based Framework*. IGI-Global.

KEY TERMS AND DEFINITIONS

Knowledge Management: Discipline in which an organization manages its knowledge in terms of resources, documents, and people skills in order to enable the adoption of insights and experiences.

Ontology: Formal representation of the knowledge in a given domain by a set of concepts and the relationships between those concepts.

OWL: Set of languages designed and established in order to be used in Semantic Web and more in detail for authoring ontologies.

Semantic Web: Evolution of the traditional Web in which contents are described in a unified manner.

Web 2.0: Set of technologies associated with WWW in which contents are usually created by users (prosumers) in contrast to websites where users (consumers) are limited to the passive viewing of content that was created for them.

This work was previously published in Customer Relationship Management and the Social and Semantic Web: Enabling Cliens Conexus, edited by Ricardo Colomo-Palacios, João Varajão and Pedro Soto-Acosta, pp. 1-22, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 29

Similarity-Based Cluster Analysis for the Cell Formation Problem

Riccardo Manzini
University of Bologna, Italy

Riccardo Accorsi
University of Bologna, Italy

Marco Bortolini
University of Bologna, Italy

ABSTRACT

This chapter illustrates the cell formation problem (CFP) supported by similarity based methods. In particular, problem oriented indices are based on several factors which play an important role in the determination of the value of similarity between two generic machines, e.g. the number of machines visited by each part, the sequence of manufacturing operations, the production quantity for each part, et cetera. A numerical example illustrates the basic steps for the implementation of an effective hierarchical procedure of clustering machines into manufacturing cells and parts/products into families of parts. Literature presents many indices, but a few significant case studies and instances not useful to properly compare them and support the best choice given an operating context, i.e. a specific production problem. As a consequence the authors illustrate an experimental analysis conducted on a literature problem oriented instance to compare the performance of different problem settings and define best practices and guidelines for professional and practitioners.

INTRODUCTION

Group technology (GT) is a manufacturing philosophy for the identification of similar parts and grouping them to take advantages from their similarities in design and manufacturing (Manzini et al. 2010). A special application of GT is *cellular manufacturing* (CM), defined as a hybrid system including the advantages of both flexible and mass production approaches. CM can be defined as an application of GT that involves grouping machines

DOI: 10.4018/978-1-4666-1945-6.ch029

based on the parts manufactured by them. The design of a CM system is called the *cell formation* (CF) problem and includes also the definition of families of "parts", i.e. products and components, assigned to the groups of manufacturing resources, called "machines".

Since 1966 when the first contribution on CM and its topics was published (Yin and Yasuda 2006), the large number of advantages presented by CM compared to batch production (generally implemented in the so-called *functional layouts* or *job shop* systems) have been widely discussed in the literature, e.g. inventory level reduction, production lead time reduction, reduced set-up times, etc. The main difference between a traditional job shop environment and a CM environment is in the grouping and layout of machines: in a job shop system, machines are grouped on the basis of their functional similarities; in a CM environment each cell is dedicated to the manufacture of a specific part family, and the machines in each cell have dissimilar functions (Heragu 1997).

An effective approach to forming manufacturing cells and introducing families of similar parts, consequently increasing production volumes and machine utilization, is the use of similarity coefficients in conjunction with clustering procedures.

Recent studies and applications on cluster analysis (CA) to industrial problems and applications are illustrated by Manzini and Bindi (2009) in transportation issues, Bindi et al. (2009) in warehousing and storage systems, Manzini et al. (2006) and (2001) in GT and CM.

Object of this chapter is the introduction, illustration and application of a cluster based systematic procedure for the design of a CM system by the adoption of *general purpose* and *problem oriented similarity* indices.

A general design of a CM system consists of the following three basic activities (Papaioannou and Wilson, 2010):

1. part families formation usually formed according to their processing requirements;
2. machine groups formation. These groups are usually called "manufacturing cells" and "clusters";
3. part families assignment to cells.

Three different strategies to execute these activities can be applied:

1. *Part family identification (PFI) strategy.* Part families are formed first and then machines grouped into families in accordance to the part families formation;
2. *Machine group identification (MGI) strategy.* Manufacturing cells are first created and then parts are allocated to cells;
3. *Part family/machine grouping (PF/MG) strategy.* Part families and manufacturing cells are formed simultaneously.

This chapter adopts the second strategy. As a consequence, this chapter illustrates a systematic procedure for the cell formation problem, i.e. the allocation of machines to cells. The number of cells to be formed is not known in advance. In a second decision step the assignment of manufacturing parts to the previously defined clusters is executed in accordance with a known processing sequence.

The simultaneous parts and machines clustering processes is usually based on the minimization of intercell movement of parts (Stawowy 2004) which specifically deals with the CF problem and methods. In other words, the object is to minimize the interactions between manufacturing cells, where an interaction occurs if a part requires machines belonging to two or more cells. The degree of interaction between manufacturing cells is measured by the number of the "exceptional elements" as illustrated below in the discussion of the efficiency in the formation process.

The remainder of this chapter is organized as follows: Section 2 presents a literature review on CM and CF problems, Section 3 illustrates the proposed similarity based hierarchical clustering process based on the application of a *threshold* *level of group similarity* as introduced by Manzini et al. (2010). In particular, Section 3 presents both a set of general purpose similarity indices and a set of problem oriented indices. Section 4 reports the most important clustering performance evaluation metrics. Section 5 illustrates a few significant numerical examples, and Section 6 discusses about the results obtained by an experimental analysis conducted on an instance of the literature adopting different settings of the decision problem. Finally Section 7 presents conclusions and further research.

LITERATURE REVIEW

A survey on CF problem methodologies is presented by Papaioannou and Wilson (2010). They distinguish three main categories: informal/visual methods, part coding analysis methods, and production based methods which can be further classified as follows (see Figure 1):

- cluster analysis (CA)
- graph partitioning approach
- mathematical programming method
- heuristic and metaheuristic algorithms
- artificial intelligent methodologies.

In particular, the CA, which is the adopted set of methods for the CF problem and CM in this chapter, groups either objects or entities into clusters such that the generic cluster is made of individual with high degree of homogeneity, i.e. “natural association”, but different clusters have very little association between them.

The proposed CA is supported by the application of “agglomerative methods” where the process of grouping starts with singleton clusters and merges them into larger sets. The systematic approach proposed in this chapter is similarity based, adopts cluster analysis and heuristic agglomerative and hierarchical algorithms.

SIMILARITY BASED HIERARCHICAL CLUSTERING PROCESS

The clustering process of parts and machines for CM presented and applied in this chapter is based on the adoption of similarity indices both “general purpose” and “problem oriented”. The main decisional steps are cited by Manzini et al. (2010) and are discussed in this section in detail. Figure 2 illustrates the proposed systematic procedure for the formation of manufacturing cells and homogeneous part families. This procedure takes inspiration from the *cluster analysis (CA)*,

Figure 1. CF solution methods classification

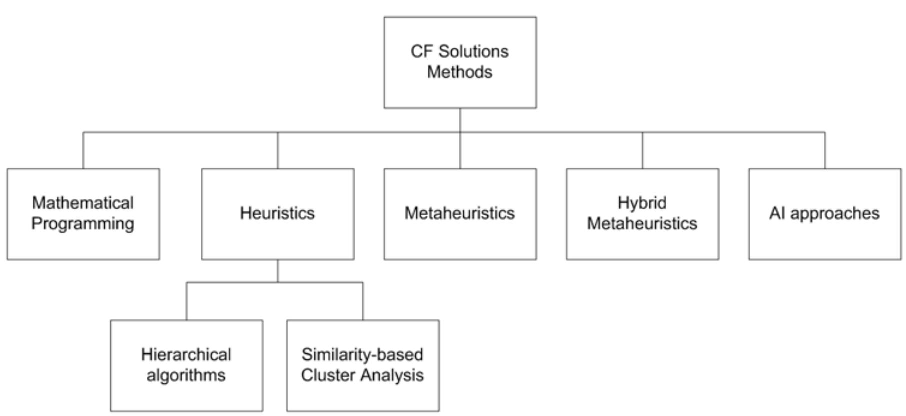
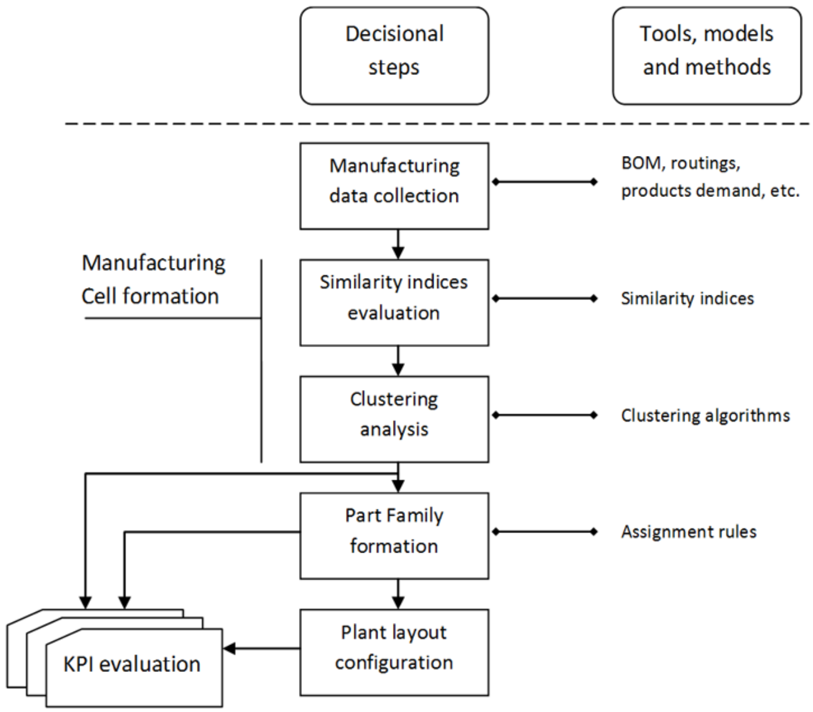


Figure 2. Systematic procedure for the similarity based hierarchical clustering



as a set of statistics based techniques and tools designed for grouping items and partitioning a set of elements in different research areas, e.g. economics, medicine, biology, etc., and is made of the following main steps.

Step 1: Manufacturing Data Collection

It deals with the analysis of product mix including all parts and components, both *product structure*, e.g. the bill of materials (BOM), and *production process*, e.g. the manufacturing work cycle, frequently called routing. The level of detail for the collection of data is significantly influenced by the adopted similarity index for the CA. In particular, in presence of a *general purpose index* the necessary data generally refer to the assignment of products and components, the so called "parts", to the manufacturing resources, the so called "machines". The adoption of a *product oriented similarity index* justifies the collection of several product data, e.g. expected production demand, manufacturing process unit time eventually including the set-up time, sequence of visited machines, existence of alternative manufacturing routines, tools assignment and availability, production costs, etc.

The basic and elementary data collection activity is usually supported by the construction of the well known *part-machine incidence matrix* (PMIM) whose generic entry a_{ik} is defined as follows:

$$a_{ik} = \begin{cases} 1 \text{ if part i visits machine k} \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where

$i=1,..,I$ part index
$k=1,..,K$ machine index.

The transpose of the part-machine matrix is known as the *machine-part incidence matrix* (MPIM).

An example of a MPIM is illustrated in Table 3. Obviously, the collection of data for the evaluation of problem oriented indices is more onerous, but the performance of the clustering process is expected to be higher than in presence of not problem oriented indices.

Step 2: Similarity Index Evaluation

The clustering activity is a process of grouping and set-partitiong items in homogeneous and disjunctive groups, called "clusters". The generic cluster of parts is called "part-family", while the cluster of machines is called "manufacturing cell". A similarity index refers to a generic pair of items of the same typology in the beginning mix of parts and machines to be partitioned. The index measures the degree to which two items, e.g. two different machines, need to belong to the same cluster, i.e. a manufacturing cell of homogeneous machines.

The literature presents a very large number of similarity coefficients. A taxonomy for these indices is presented by Yin and Yasuda (2006). They distinguish the previously cited two distinct main groups of coefficients: *problem-oriented* (*l*1.1 in Figure 3) and *general purpose* (*l*1.2). Problem-oriented measures are specifically designed for application to manufacturing problems in industry, while general purpose are used in many disciplines, e.g. medicine, sociology, biology, economics, decision science, etc.

Table 1 lists a not exhaustive set of general purpose indices S_{ij} as proposed by the literature and based on the following assumptions: given two machines i and j, a is the number of parts visiting both machines, b is the number of parts visiting machine i but not j, c is the number of parts visiting machine j but not i, d is the number of parts visiting neither machine i, j. These coefficients are applied in the numerical example and experimental analysis illustrated in this chapter.

Jaccard - J is the most commonly used general-purpose similarity coefficient in the literature. Jaccard similarity coefficient between machine i and machine j is defined as follows (McAuley, 1972):

$$S_{ij} = \frac{a}{a+b+c}, \quad 0 \leq S_{ij} \leq 1 \qquad (2)$$

Figure 3. Similarity coefficients by Yin and Yasuda (2006)

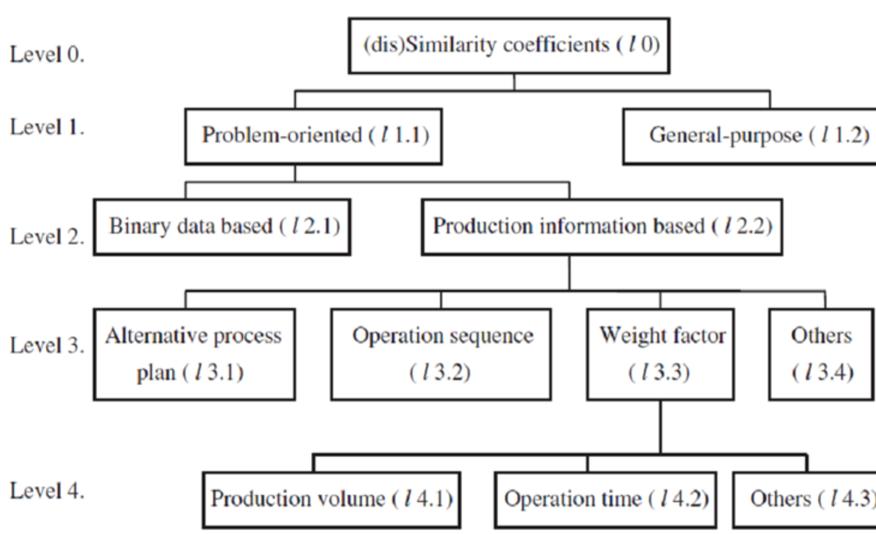


Table 1. General purpose similarity indices

Code	Coefficient	Range	S_{ij}
B	Baroni-Urbani and Buser	0–1	$[a+(ad)^{1/2}]/[a+b+c+(ad)^{1/2}]$
H	Hamann	-1 to 1	[(a+d)-(b+c)]/[(a+d)+(b+c)]
J	Jaccard	0–1	a/(a+b+c)
O	Ochiai	0–1	$a/[(a+b)(a+c)^{1/2}]$
R	Rogers and Tanimoto	0–1	(a+d)/[a+2(b+c)+d]
RM	Sarker and Islam (Relative matching)	-1 to 1	$[a+(ad)^{1/2}]/[a+b+c+d+(ad)^{1/2}]$
SK	Sokal and Sneath	0–1	2(a+d)/[2(a+d)+b+c]
SO	Sorenson	0–1	2a/(2a+b+c)
SI	Sokal and Michener (Simple matching)	0-1	(a+d)/(a+b+c+d)
RR	Russel and Rao	0-1	a/(a+b+c+d)

Problem oriented similarity coefficients can be classified into binary data based (*l*2.1 in Figure 3) and production information based (*l*2.2) similarity coefficients. The similarity coefficients at the level *l*2.1 only consider assignment information (a part need or need not a machine). Yin and Yamada (2006) present a list of binary data problem oriented indices. The level 3 (*l*3), see Figure 3, introduces different manufacturing factors for cell formation, e.g. machine requirement, machine setup, utilization, workload, alternative routings, machine capacities, operation sequences, set-up cost and cell layout. Finally level 4 (*l*4) introduces different weights for the evaluation of weighted similarity coefficients in order to adjust the strength of matches or misses between objects pairs to reflect the resemblance value more realistically (Yin and Yamada, 2006).

A few examples of problem oriented measures are reported below and are based on the following basic assumptions: n is the number of parts, $k=1,\ldots,n$ the generic part, m the number of machines and $i,j = 1,\ldots, m$ the generic machine (Alhourani and Seifoddini 2007). Then:

$$x_{ijk} = \begin{cases} 1 \text{ if part } k \text{ visits both machine } i \text{ and machine } j \\ 0 \end{cases}$$

$$y_{ijk} = \begin{cases} 1 \text{ if part } k \text{ visits either machine } i \text{ or machine } j \\ 0 \end{cases}$$

- S - Seifoddini (1987). He modifies the Jaccard's coefficient incorporating the production volume as follows:

$$S_{ij} = \frac{\sum_{k=1}^{n} x_{ijk} \cdot m_k}{\sum_{k=1}^{n} y_{ijk} \cdot m_k + \sum_{k=1}^{n} x_{ijk} \cdot m_k} \qquad (3)$$

where m_k planned production volume during a predefined period of time for part type *k.*

- GS - Gupta and Seifoddini (1990). They extend the Jaccard's similarity coefficient to incorporate the effect of the operational time, operational sequence and production volume. The proposed index is:

$$S_{ij} = \frac{\sum_{k=1}^{n} (x_{ijk} \cdot T_{ijk} + \sum_{o=1}^{n_k} z_{ko}) \cdot m_k}{\sum_{k=1}^{n} (x_{ijk} \cdot T_{ijk} + \sum_{o=1}^{n_k} z_{ko} + y_{ijk}) \cdot m_k} \qquad (4)$$

where n_k number of times the part type *k* visits both machines in row, i.e. sequentially

$$z_{ko} = \begin{cases} 1 \text{ if part } k \text{ visits both machines } i \text{ and } j \text{ in row} \\ 0 \end{cases}$$

$$T_{ijk} = \frac{\min\left\{T_{ik}, T_{jk}\right\}}{\max\left\{T_{ik}, T_{jk}\right\}} \quad (5)$$

T_{ik} global time spent by part type *k* on machine *i*

T_{jk} global time spent by part type *k* on machine *j*.

- SH - Seifoddini and Hsu (1994). They propose a weighted similarity coefficient with the aim of eliminating *improper machine assignment* by giving a higher weight to components having common operations on both the machines.

$$S_{i,j} = \frac{\sum_{k=1}^{n} f_{bk} \cdot x_{ijk}}{\sum_{k=1}^{n} f_{bk} \cdot x_{ijk} + \sum_{k=1}^{n} f_{ek} \cdot y_{ijk}} \quad (6)$$

where

f_{bk} weighting factor for parts visiting both machines *i* and *j*

f_{ek} weighting factor for parts visiting either machine *i* or *j* but not both.

In the numerical example and experimental analysis illustrated below, the adopted value of f_{bk} and f_{ek} are respectively 0.6 and 0.4.

- N - Nair and Narendran (1998).They propose a similarity measure as the ratio of the sum of the movements common to the machines *i* and *l*, and the sum of the total number of movements to and from machines *i* and *l*.

$$s_{il} = \frac{c_i + c_l}{t_i + t_l} \quad (7)$$

$$t_i = \sum_{j=1}^{n} \sum_{p=1}^{n_{ji}} w_j t_{jip} \quad (8)$$

$$c_{il} = \sum_{j=1}^{n} \sum_{p=1}^{n_{jil}} w_j c_{jilp} \quad (9)$$

where

m number of machines

n number of parts

n_{ji} number of times part *j* visits the *i*th machine

n_{jil} number of times part *j* visits the *i*th and *j*th machine

b_{jip} operation sequence number if the part *j* visits machine *i* for *p*th time ($0 \le p \le n_{ji}$); zero otherwise

b_{jlp} operation sequence number if the part *j* visits machine l for *p*th time ($0 \le p \le n_{ji}$); zero otherwise

r_j maximum number of operations for part *j*

$$t_{jip} = \begin{cases} 0 \ \text{if} \ b_{jip} = 0 \\ 1 \ \text{if} \ b_{jip} = 1 \ \text{or} \ r_j \\ 2 \text{ otherwise} \end{cases}$$
$$c_{jilp} = \begin{cases} 0 \ \text{if} \ b_{jip} = 0 \ \text{or} \ b_{jlp} = 0 \\ 1 \ \text{if} \ b_{jip} = 1 \ \text{or} \ r_j \\ 2 \text{ otherwise} \end{cases} \quad (10)$$

Equation (8) accounts for the total number of movements to and from the machine *i*. Equation (9) quantifies the total number of movements to and from machine *i* made by all parts that visit both machine *i* and machine *l*.

Step 3: Clustering Analysis and Manufacturing Cells Formation

The Clustering step deals with the grouping of machines into different homogeneous clusters so that machines in each cluster have high values of correlation. The clustering analysis is supported by the application of a grouping algorithm, which is a hierarchical heuristic for the partitioning of machines into disjunctive cells. There are usually substantial differences between the machine groups obtained by clustering, but the individuals within an individual machine group are similar because they are similarly visited by different parts/components. The clustering process is generally supported by one of the following well known hierarchical algorithms (Mosier 1989): Complete Linkage Method (CLINK) also known as farthest neighbor (fn) clustering, Single Linkage Method (SLINK) also known as nearest neighbor (*nn*) clustering, Unweighted Pair-Group Method using Arithmetic Average (UPGMA), Weighted Pair-Group Method using Arithmetic Average (Average Linkage), and Unweighted Pair-Group Method using Centroid (UPGMC).

In particular, the authors (of this chapter) choose to illustrate the *fn* clustering and the *nn* clustering algorithm (Aldenderfer and Blashfield, 1984). The generic algorithm is based on a scheme that erases rows and columns in a similarity matrix *S* that collects the values of similarity for each couple of items, as old clusters are merged into new ones by the degradation of the level of similarity within each under construction cluster. The similarity matrix *S* (dimension $N x N$) contains all the correlations among the machines $s(i,j)$, calculated as described in the previous step. To the clusterings are assigned sequence numbers $0, 1, \ldots, (n-1)$, and $L(k)$ is the level of similarity of the *k-th* clustering. A cluster with sequence number *m* is denoted *(m)* and the correlation between clusters *(r)* and *(s)* is denoted $s[(r),(s)]$. In particular, the *nn* algorithm is composed of the following steps:

1. Begin with the disjoint clustering having level $L(0) = 1$ and sequence number $m = 0$.
2. Find the least dissimilar pair of clusters in the current clustering, say pair *(r)* and *(s)*, according to:

$$s\left[\left(r\right),\left(s\right)\right] = \max_{i,j}\left\{s\left[\left(i\right),\left(j\right)\right]\right\} \tag{11}$$

where the maximum is over all pairs of clusters in the current clustering.

3. Increment the sequence number: $m = m + 1$. Merge the groups of items *(r)* and *(s)* into a single cluster to form the next clustering *m*. Set the level of this clustering to:

$$L(m)=s[(r),(s)] \tag{12}$$

4. Update the similarity matrix *S* by deleting the rows and columns corresponding to clusters *(r)* and *(s)* and adding a row and column corresponding to the newly formed cluster. The level of correlation between the new cluster, denoted *(r,s)* and the old generic cluster *(k)* is defined as:

$$s[(k), (r,s)] = \max \{s[(k),(r)], s[(k),(s)]\} \tag{13}$$

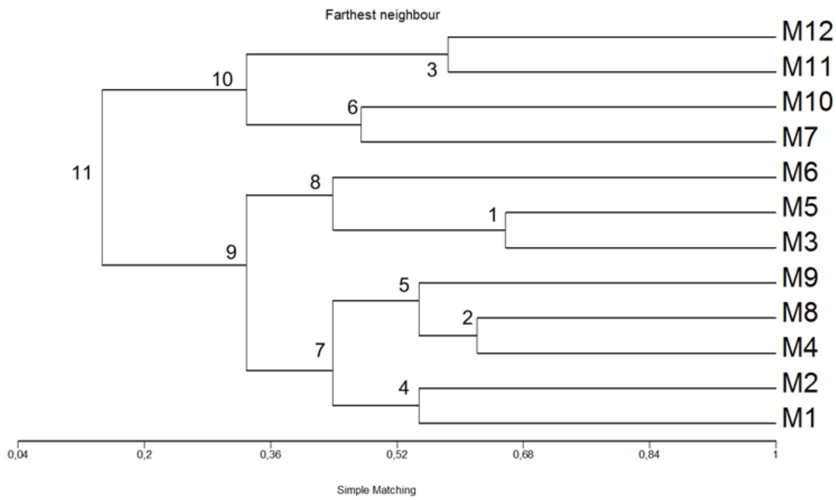
If all the machines are in one cluster, stop. Else, go to point 2.

The *fn* algorithm is based on the same scheme and equation (13) in point 4 is modified as follows:

$$s[(k), (r,s)] = \min \{s[(k),(r)], s[(k),(s)]\} \tag{14}$$

The dendrogram is the graphical representation of the process of degradation of the similarity level as the result of the grouping process executed by the clustering algorithm (Sokal and Sneath 1968, McAuley 1972). Figure 4 exemplifies a dendrogram generated by the execution of the clustering process to an instance of example.

Figure 4. Dendrogram: Simple Matching and farthest neighbour, [1-11] nodes



As a consequence the generic algorithm classifies the machines on the basis of the similarity of the manufacturing characteristics and hierarchically executes aggregations of machines into clusters reducing progressively the level of similarity, i.e. homogeneity, within the under construction heterogeneous clusters.

The result of the clustering process does not only depend on the rule adopting for grouping but also on the threshold level of similarity adopted for grouping, i.e. the minimal admissible value of similarity within the generic cluster as discussed by Manzini et al. (2010). In particular, they demonstrate that a best performing percentile based level of threshold similarity exists to optimize the cell formation process in presence of general purpose similarity metrics. The percentile based threshold value is a range of group similarity measurements which cuts the dendrogram at the percentile number of aggregations identified by the clustering rule, as follows:

$$T\_value_{\%_p} \in \left]simil\left\{\left\lfloor \%_p \times N_{nodes} \right\rfloor\right\}, simil\left\{\left\lceil \%_p \times N_{nodes} \right\rceil\right\}\right[\quad (15)$$

where

$\%_p$ percentile of aggregations, expressed as a percentage

N_{nodes} number of nodes generated by the clustering algorithm

Simil$\{N\}$ similarity value which corresponds to the node *N*.

The basic idea of this criterion is properly illustrated by the example discussed in Section 5. Figure 4 presents a dendrogram generated by the application of clustering analysis. *m1,..,m12* are the identifications of machine items. The number within the diagram (1,..,12) identify the aggregations (called nodes) ordered in agreement with the similarity measurements. In particular, low numbers identify aggregations between under construction clusters characterized by a high level of similarity (very similar clusters). Table 5 reports the list and configuration of nodes as generated by the application of the clustering algorithm in agreement with Figure 4. Assuming the 80° percentile of aggregation (i.e. $\%_p=0.80$) as follows (see Equation (15)):

$$T\_value_{80°} \in \left]simil\left\{\left\lceil 0.80 \times 12 \right\rceil\right\}, simil\left\{\left\lfloor 0.80 \times 12 \right\rfloor\right\}\right[= \\ = \left]simil\left\{10\right\}, simil\left\{9\right\}\right[= \left]0.329,\ 0.329\right[= 0.329$$

Similarly, for a percentile value of 20° (i.e. $\%_p = 0.20$):

$$T\_value_{20°} \in \left]simil\left\{\left\lceil 0.2 \times 12 \right\rceil\right\}, simil\left\{\left\lfloor 0.2 \times 12 \right\rfloor\right\}\right[= \\ = \left]simil\left\{3\right\}, simil\left\{2\right\}\right[= \left]0.585,\ 0.622\right[$$

Step 4: Part Families Formation

Given a generic solution for cells, a part may have to visit more than one group of machines before it is completed. Consequently, the generic part has to be assigned to the manufacturing cell with the minimum number of inter-group journeys. Another way of reducing the number of inter-group journeys is to duplicate machines, but it can be very expensive in terms of space and monetary costs.

Part families can be formed concurrently with the cell/group machine formation (CF steps illustrated above), or otherwise executed after the cells have been defined. In particular, the second of these hypotheses is adopted in this paper, and in particular a hierarchical heuristic rule is applied to assign parts to manufacturing cells. The main steps are:

- Given a configuration of the disjunctive groups of machines (i.e. manufacturing cells) name them as $c=1,..,C$. Then quantify the following measurements for each part *i* and each manufacturing cell *c* in accordance with the working cycle of part *i*:
    - number of intra-cell movements: ICM_{ic};
    - number of tasks executed in the cell: $NTask_{ic}$;
    - processing time of *i* in *c*: $Time_{ic}$.
- Assign part *i* to cell *c** where:

$$ICM_{ic^*} > \max_{\substack{c=1,..,C \\ c \neq c^*}} \left\{ICM_{ic}\right\} \qquad (16)$$

If *c** does not exist than GO to STEP III.

- Assign part *i* to cell *c'* shown in Equation 17 (Box 1). If *c'* does not exist than GO to STEP IV.
- Assign part *i* to cell *c''* shown in Equation 18 (Box 1). If c'' does not exist than assign part *i* randomly to a cell in the last equation shown in Box 1.

Box 1.

$$NTask_{ic'} > \max_{\substack{c=1,..,C \\ c \neq c'}} \left\{ NTask_{ic} : ICM_{ic} = \max_{\substack{c^*=1,..,C \\ c^* \neq c}} \left\{ICM_{ic^*}\right\} \right\} \qquad (17)$$

$$Time_{ic''} > \max_{\substack{c=1,..,C \\ c \neq c''}} \left\{ Time_{ic} : ICM_{ic} = \max_{\substack{c^*=1,..,C \\ c^* \neq c}} \left\{ICM_{ic^*}\right\} \wedge NTask_{ic} = \max_{\substack{c'=1,..,C \\ c' \neq c}} \left\{Ntask_{ic'}\right\} \right\} \qquad (18)$$

$$\left\{ c = 1,\dots,C : ICM_{ic} = \max_{\substack{c^*=1,..,C \\ c^* \neq c}} \left\{ICM_{ic^*}\right\} \wedge NTask_{ic} = \max_{\substack{c'=1,..,C \\ c' \neq c}} \left\{NTask_{ic'}\right\} \wedge Time_{ic} = \max_{\substack{c''=1,..,C \\ c'' \neq c}} \left\{Time_{ic''}\right\} \right\}$$

A result of the assignment of parts to the manufacturing cells is the so-called *block-diagonal incidence matrix* shown in Figure 5 as the result of the application of the proposed procedure to the instance illustrated in Section 5.

Step 5: Plant Layout Configuration

This step deals with the determination of the location of each manufacturing resource (machines and human resources) in the production area. Layout decisions are significantly influenced by the configuration of cells and part families in CM systems, but they are omitted in this chapter. Nevertheless a few significant performance measures on layout results are quantified as clearly explained below.

CLUSTERING PERFORMANCE EVALUATION

Sarker (2001) presents, discusses, and compares the most notable measurements of grouping efficiency in CM. The measurements adopted in the following illustrated experimental analysis are based on the following definitions:

- *lock*. This is a submatrix of the machine-part matrix composed of rows representing a part family and columns representing the related machine cell.
- *Void*. This is a "zero" element appearing in a diagonal block (see Figure 4).
- *Exceptional element*. This is a "one" appearing in off-diagonal blocks (see Figure 4). The exceptional element causes inter-cell movements.

A set of CM measurements of performance quantified in the proposed experimental analysis is now reported and discussed. (high) and (low) labels refer to the expected values for best performing the CF and CM.

Problem Density: PD

$$PD = \frac{\text{number of "ones" in the incidence matrix}}{\text{number of elements in the incidence matrix}} \tag{19}$$

Global Inside cells density: ICD (high)

$$ICD = \frac{\text{number of "ones" in diagonal blocks}}{\text{number of elements in diagonal blocks}} \tag{20}$$

Figure 5. Block-diagonal matrix. Simple matching, farthest neighbour.& 75° percentile.

	p11	p12	p16	p17	p5	p6	p13	p14	p15	p18	p19	p9	p10	p1	p2	p3	p4	p7	p8
Cell 1																			
m11		1	1	1			1	1	1										
m12	1	1	1	1			1												
Cell 2																			
m7		1		1	1	1	1	1		1	1				1	1	1		
m10					1	1	1	1	1	1	1								
Cell 3																			
m6	1				1	1				1		1	1		1			1	1
m3												1	1						1
m5												1							1
Cell 4																			
m1					1									1	1	1	1		
m2															1	1			1
m9					1	1						1		1		1	1	1	1
m4												1	1	1	1	1	1	1	1
m8						1						1	1	1	1	1		1	1

Exceptional elements: EE (low)

EE = Number of exceptional elements in the off-diagonal blocks (21)

Ratio of non-zero elements in cells: REC

$$REC = \frac{\text{total number of "ones"}}{\text{number of elements in diagonal blocks}} \quad (22)$$

Grouping Efficiency: η (Sarker 2001) (high)

It is a weighted average of two functions and it is defined as:

$$\eta = q\eta_1 + (1-q)\eta_2$$
$$\eta_1 = \frac{e_d}{\sum_{r=1}^{k} M_r N_r} \quad (23)$$
$$\eta_2 = 1 - \frac{e_o}{mn - \sum_{r=1}^{k} M_r N_r}$$

where

e_d number of 1s in the diagonal blocks
e_o number of 1s in the off-diagonal blocks
k number of diagonal blocks
M_r number of machines in the *r*th cell
N_r number of components in the *r*th part-family
q weighting factor (0≤q≤1) that fixes the relative importance between voids and inter-cell movements. If q=0.5 both get the same importance: this is the value adopted in the numerical example and in the experimental analysis illustrated in sections 5 and 6.

Quality Index: QI (Seifoddini and Djassemi 1994, 1996) (high)

It is a measure of independence of machine-component groups. High values of QI are expected in presence of high independency. QI is defined as:

$$QI = 1 - \frac{ICW}{PW} \quad (24)$$

where

ICW total intercellular workload
PW total plant workload.
ICW and PW can be defined as:

$$ICW = \sum_{c=1}^{C}\sum_{i=1}^{m}\left[Y_{ic}\left(\sum_{k=1}^{n}(1-Z_{kc})X_{ik}m_k T_{ik}\right)\right] \quad (25)$$

$$PW = \sum_{i=1}^{m}\sum_{k=1}^{n} X_{ik} m_k T_{ik} \quad (26)$$

where *n* is the number of parts, *k=1,…,n* the generic part, *m* the number of machines and *i,j=1,…,m* the generic machine. This is the notation previously introduced

$$Y_{ic} = \begin{cases} 1 \text{ if machine } i \text{ is assigned to cell } c \\ 0 \text{ otherwiswe} \end{cases}$$
$$Z_{kc} = \begin{cases} 1 \text{ if part } k \text{ is assigned to cell } c \\ 0 \text{ otherwiswe} \end{cases}$$
$$X_{ki} = \begin{cases} 1 \text{ if part } k \text{ has operation on machine } i \\ 0 \text{ otherwiswe} \end{cases}$$

m_k volume of part *k*
T_{ki} processing time of part *k* on machine *i*

QI measures the number of intracellular movements which ask to be maximized minimizing intercellular ones.

Now the authors introduce a new grouping efficiency based on QI as previously defined.

Grouping Efficiency based on QI: ηQI (high)

$$\eta_{QI}=q\eta_1+(1-q)QI \tag{27}$$

The adopted value of weighting factor *q* is:

$$q=\frac{\sum_{r=1}^{k}M_rN_r}{mn} \tag{28}$$

Grouping Efficacy: τ (Kumar and Chandrasekharan 1990) (high)

Group efficacy can be quantified by the application of the following equation:

$$\tau=\frac{e-e_0}{e+e_v} \tag{29}$$

where

e total number of "ones" in the matrix (i.e. the total number of operations)

$e_0=EE$ number of exceptional elements (number of "ones" in the off-diagonal blocks)

e_v number of voids (number of "zeros" in the diagonal blocks).

Grouping measure: ηG (Miltenburg and Zhang 1991) (high)

It gives higher values if both the number of voids and exceptional elements are fewer, and it is defined as:

$$\begin{aligned}\eta_g&=\eta_u-\eta_m\\ \eta_u&=\frac{e_1}{e_1+e_v}\\ \eta_m&=\frac{e_o}{e}\end{aligned} \tag{30}$$

where

η_u ratio of the number of 1s to the number of total elements in the diagonal block (this is the inside cell density - ICD)

η_m ratio of exceptional elements to the total number of 1s in the matrix

e_1 number of 1s in the diagonal block.

Group technology efficiency: GTE (Nair and Narendran 1998) (high)

It is defined as the ratio of the difference between the maximum number of inter-cell travels possible and the number of inter-cell travels actually required to the maximum number of inter-cell possible:

$$\begin{aligned}GTE&=\frac{I-U}{I}\\ I&=\sum_{j=1}^{n}(r_j-1)\\ U&=\sum_{j=1}^{n}\sum_{s=1}^{r_j-1}xl_{js}\end{aligned} \tag{31}$$

where

I maximum number of inter-cell travels

U number of inter-cell movements required by the system

r_j maximum number of operations for component j

$$xl_{js}=\begin{cases}0 \text{ if operations } s,\ s+1 \text{ are performed in the same cell}\\ 1 \text{ otherwise}\end{cases}$$

Bond efficiency: BE (high)

This is an important index because depends on both the within-cell compactness (by the ICD) and the minimization of inter-cell movements by the GTE. It is defined as:

$$BE = q \cdot IDC + (1-q) \cdot GTE \quad (32)$$

The adopted value of weight q in the experimental analysis is 0.5.

NUMERICAL EXAMPLE

This section presents a numerical example which relates to a problem oriented instance presented by De Witte (1980) and made of 19 parts and 12 machines. Manufacturing input data are reported in Table 2.

Table 3 reports the *12x19 machine-part* incidence matrix useful for the evaluation of a general purpose similarity index.

A General Purpose Evaluation

This section presents the results obtained by the application of a general purpose similarity index in cluster analysis for the cell formation problem.

Table 4 reports the result of the evaluation of the general purpose index known as Simple Matching (SI) and defined in Table 1. Figure 4 shows the dendrogram generated by the application of the fn combined with the SI similarity coefficient. In particular a sequence of numbers is explicitly reported in figure for each node of the diagram. The generic node corresponds to a specific aggregation ordered in agreement with the similarity metric and the adopted hierarchical rule. The list of nodes and aggregations, the related values of similarity, and the number of objects per group are also reported in Table 5. The obtained number of nodes is 11.

Now it is possible to define a partitioning of the available set of machines by the identification of a cut value, the so called "cutting threshold similarity value". The adopted level of homogeneity within the generic cluster is the percentile-based threshold measure discussed in Section 3.

Given the dendrogram in Figure 4 and assuming a threshold percentile cut value equal to 20°, the corresponding range of similarity is (0.585, 0.622) as demonstrated in Section 3. The obtained configuration of the manufacturing cells (nine different cells are obtained) is:

Cell 1 (single machine): M12
*Cell 2 (*single *machine): M11*

Table 2. Manufacturing input data, De Witte (1980)

Part	Volume	Work Cycle	Processing Time
p1	2	m1, m4, m8, m9	20, 15, 10, 10
p2	3	m1, m2, m6, m4, m8, m7	20, 20, 15, 15, 10, 25
p3	1	m1, m2, m4, m7, m8, m9	20, 20, 15, 25, 10, 15
p4	3	m1, m4, m7, m9	20, 15, 25, 15
p5	2	m1, m6, m10, m7, m9	20, 15, 20, 25, 15
p6	1	m6, m10, m7, m8, m9	15, 50, 25, 10, 15
p7	2	m6, m4, m8, m9	15, 15, 10, 15
p8	1	m3, m5, m2, m6, m4, m8, m9	30, 50, 20, 15, 15, 10, 15
p9	1	m3, m5, m6, m4, m8, m9	30, 50, 15, 15, 10, 15
p10	2	m3, m6, m4, m8	30, 15, 15, 10
p11	3	m6, m12	15, 20
p12	1	m11, m7, m12	40, 25, 20
p13	1	m11, m10, m7, m12	40, 50, 25, 20
p14	3	m11, m7, m10	40, 25m 50
p15	1	m11, m10	40, 50
p16	2	m11, m12	40, 20
p17	1	m11, m7, m12	40, 25m 20
p18	3	m6, m7, m10	15, 25, 50
p19	2	m10, m7	50, 25

Table 3. Machine-part incidence matrix

	p1	**p2**	**p3**	**p4**	**p5**	**p6**	**p7**	**p8**	**p9**	**p10**	**p11**	**p12**	**p13**	**p14**	**p15**	**p16**	**p17**	**p18**	**p19**
m1	1	1	1	1	1														
m2		1	1					1											
m3								1	1	1									
m4	1	1	1	1			1	1	1	1									
m5								1	1										
m6		1			1	1	1	1	1	1	1							1	
m7		1	1	1	1	1						1	1	1			1	1	1
m8	1	1	1			1	1	1	1	1									
m9	1		1	1	1	1	1	1	1										
m10					1	1							1	1	1			1	1
m11												1	1	1	1	1	1		
m12											1	1	1			1	1		

*Cell 3 (*single *machine): M10*
Cell 4 (single machine): M7
*Cell 5 (*two *machines): M3, M5*
Cell 6 (single machine): M9
Cell 7 (two machines*): M8, M4*
*Cell 8 (*single *machine): M2*
Cell 9 (single machine): M1

In case a cut value corresponds to one or more nodes generated by the hierarchical process of aggregation, it is possible to include (exclude) the node in the formation of cells. In particular, assuming a level of threshold similarity equal to 80°, two alternative configurations can be obtained as the result of inclusion/exclusion of one or more nodes of the dendrogram located in correspondence of the cutting level:

Case 1: Including node 10 and node 9

Table 4. Simple matching similarity matrix

	m1	**m2**	**m3**	**m4**	**m5**	**m6**	**m7**	**m8**	**m9**	**m10**	**m11**	**m12**
m1	1.0000											
m2	0.5479	1.0000										
m3	0.4021	0.5479	1.0000									
m4	0.5118	0.5118	0.5118	1.0000								
m5	0.4389	0.5847	0.6576	0.4750	1.0000							
m6	0.3292	0.4021	0.4750	0.4389	0.4389	1.0000						
m7	0.4021	0.3292	0.1826	0.2195	0.2195	0.2556	1.0000					
m8	0.4389	0.5118	0.5118	0.6215	0.4750	0.5118	0.2194	1.0000				
m9	0.5118	0.4389	0.4389	0.5479	0.4750	0.4389	0.2924	0.5479	1.0000			
m10	0.3292	0.3292	0.3292	0.1465	0.3653	0.3292	0.4750	0.2194	0.2924	1.0000		
m11	0.2924	0.3653	0.3653	0.1826	0.4021	0.1465	0.3653	0.1826	0.1826	0.4389	1.0000	
m12	0.3292	0.4021	0.4021	0.2194	0.4389	0.2556	0.3292	0.214	0.2194	0.3292	0.5847	1.0000

Table 5. List and configuration of nodes generated by fn rule & SI similarity coefficient

Node	Group 1	Group 2	Simil.	Objects in Group
1	M3	M5	0.658	2
2	M4	M8	0.622	2
3	M11	M12	0.585	2
4	M1	M2	0.548	2
5	Node 2	M9	0.548	3
6	M7	M10	0.475	2
7	Node 4	Node 5	0.439	5
8	Node 1	M6	0.439	3
9	Node 7	Node8	0.329	8
10	Node 6	Node 3	0.329	4
11	Node 9	Node 10	0.146	12

Cell 1 (four machines): M12, M11, M10 and M7
Cell 2 (eight machines): M6, M5, M3, M9, M8, M4, M2, M1.

Case 2: Not including node 10 and node 9
Cell 1 (two machines): M12, M11
Cell 2 (two machines): M10, M7
Cell 3 (3 machines): M6, M5, M3
Cell 4 (5 machines): M9, M8, M4, M2, M1.

The second column in Table 8 reports the obtained values of the performance evaluation for the case study object of this numerical example adopting the Simple Matching similarity index, the fn heuristic, and the cutting threshold percentile value equal to 75°.

A Problem Oriented Evaluation

Table 6 reports the result of the evaluation of the problem oriented similarity coefficient as proposed by Nair and Narendran (1998). Figure 6 shows the dendrogram as the result of the application of the fn clustering rule. The generic node of the dendrogram corresponds to a specific aggregation ordered in agreement with the adopted similarity metric and the adopted hierarchical rule. The list of nodes and aggregations, the related values of similarity, and the number of objects in group are also reported in Table 7 as the result of the application of the fn rule and Nair and Narendran (1998) problem oriented similarity coefficient. The obtained number of nodes is 11.

Figure 6 reports the dendrogram obtained by the application of the fn clustering heuristic rule and the "Nair and Narendran" problem oriented similarity coefficient to the literature instance of interest.

Table 6. Nair & Narendran similarity matrix

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12
m1	1.0000											
m2	0.5000	1.0000										
m3	0.0000	0.2220	1.0000									
m4	0.6920	0.5000	0.4210	1.0000								
m5	0.0000	0.2860	0.6670	0.2350	1.0000							
m6	0.3450	0.3480	0.3640	0.5450	0.2000	1.0000						
m7	0.5620	0.3080	0.0000	0.3890	0.0000	0.4620	1.0000					
m8	0.5000	0.5560	0.4710	0.8570	0.2670	0.6450	0.2940	1.0000				
m9	0.6670	0.2220	0.2350	0.7150	0.2670	0.4520	0.4720	0.6150	1.0000			
m10	0.1670	0.0000	0.0000	0.0000	0.0000	0.3870	0.7060	0.0770	0.2310	1.0000		
m11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4000	0.0000	0.0000	0.4550	1.0000	
m12	0.0000	0.0000	0.0000	0.0000	0.0000	0.2310	0.2070	0.0000	0.0000	0.0950	0.5880	1.0000

Figure 6. Dendrogram by the application of Nair & Narendran similarity coefficient and the farthest neighbour

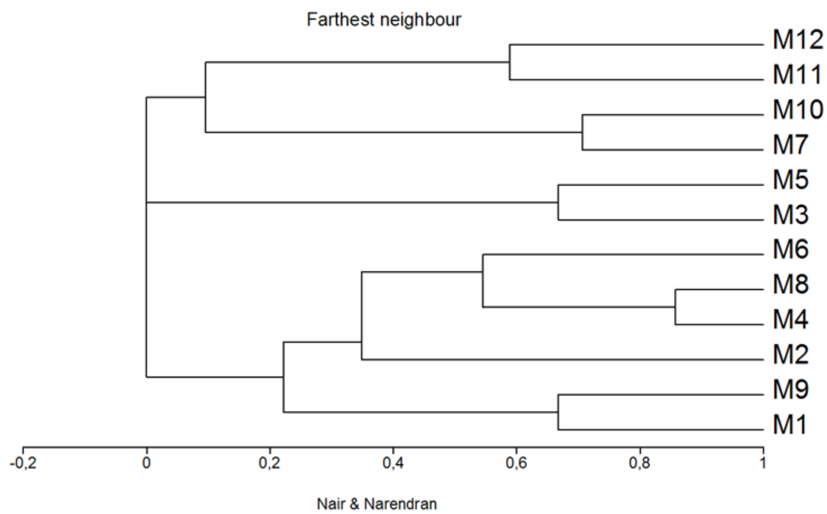


Table 7. List and configuration of nodes generated by the fn rule & Nair and Narendran (1998) similarity coefficient

Node	Group 1	Group 2	Simil.	Objects in Group
1	M4	M8	0.857	2
2	M7	M10	0.706	2
3	M1	M9	0.667	2
4	M3	M5	0.667	2
5	M11	M12	0.588	2
6	Node 1	M6	0.545	3
7	M2	Node 6	0.348	4
8	Node 3	Node 7	0.222	6
9	Node 2	Node 5	0.095	4
10	Node 8	Node 4	0	8
11	Node 10	Node 9	0	12

- Assuming $\%_p$=80°:

$$T\_value_{80^\circ} \in \left]simil\left\{\left\lceil 0.80 \times 11 \right\rceil\right\}, simil\left\{\left\lfloor 0.80 \times 11 \right\rfloor\right\}\right[= \\ = \left]simil\left\{9\right\}, simil\left\{8\right\}\right[= \left]0.095,\ 0.222\right[$$

The obtained configuration of the manufacturing cells (four different cells are obtained) is:

Cell 1 (two machines): M11, M12
Cell 2 (two machines): M7, M10
Cell 3 (two machines): M3, M5
Cell 4 (six machines): M6, M8, M4, M2, M9, M1

- Assuming $\%_p$=20°:

$$T\_value_{20^\circ} \in \left]simil\left\{\left\lceil 0.20 \times 11 \right\rceil\right\}, simil\left\{\left\lfloor 0.20 \times 11 \right\rfloor\right\}\right[= \\ = \left]simil\left\{3\right\}, simil\left\{2\right\}\right[= \left]0.667,\ 0.706\right[$$

The obtained configuration of the manufacturing cells (eleven different cells are obtained) is:

Single machine cells: Cell 1(M12), Cell 2(M11), Cell 4(M5), Cell 5(M3), Cell 6(M3), Cell7(M6), Cell 9(M2), Cell 10(M9), Cell 11(M1)
Double machines cells: Cell 3 (M7, M10), Cell 8 (M8, M4).

Table 8. Performance evaluation of numerical example; 75° percentile

Similarity index	ID	Simple matching	Nair & Narendran
Problem Density	PD	0.329	0.329
Inside Cells Density	ICD	0.705	0.828
REC	REC	0.962	1.293
Exceptional Element	EE	20	27
Grouping Efficiency [%]	n	60.2	57.9
Grouping Efficiency QI [%]	n_{QI}	68.8	72.8
Group Technology Efficiency [%]	GTE	61.9	45.5
Bond Efficiency [%]	BE	66.2	66.1
Group Efficacy [%]	τ	82.9	84.5
Grouping measure	n_G	0.438	0.468

The third column in Table 8 reports the obtained values of the performance evaluation for the case study object of this numerical example adopting the “Nair and Narendran” similarity index, the fn heuristic, and the cutting threshold percentile value equal to 75°.

Which is the best similarity index? It is not correct to try to reply to this question as is, because previous sections demonstrate that there are different factors affecting the performance of the system configuration: the similarity index, the clustering rule, the threshold cutting value of similarity, and the part assignment rule. As a consequence it is useful to measure the simultaneous effects generated by different combinations of these critical factors. Next section presents an experimental analysis conducted on the instance proposed by De Witte (1980) comparing the performance obtained adopting general purpose and problem oriented similarity metrics.

EXPERIMENTAL ANALYSIS

This section presents the results obtained by the application of the proposed systematic procedure to cell formation and parts assignment to cells (*part family formation*), as the result of different settings of the similarity and hierarchical procedure as illustrated in previous sections. This what-if analysis is applied to the problem oriented instance introduced by De Witte (1980) and reported in Table 2. This analysis represents the first step to identify the best combination of

Table 9. What-if analysis, factors and levels

	general purpouse	problem oriented
Similarity Coefficient	J, SI, H, B, SO, R, SK, O, RM, RR	S, GS, SH (f_{bk}=0.6;f_{ek}=0.4), N
Rule	CLINK, ALINK, SLINK	
Percentile	10°, 25°, 40°, 50°, 75°	

Figure 7. Block-diagonal matrix. Nair and Narendran, farthest neighbour.& 75° percentile.

	p13	p14	p15	p18	p19	p11	p12	p16	p17	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
Cell 1																			
m7	1	1		1	1		1		1		1	1	1	1	1				
m10	1	1	1	1	1									1	1				
m3																	1	1	1
m5																	1	1	
Cell 2																			
m11	1	1	1				1	1	1										
m12	1					1	1	1	1										
Cell 3																			
m1										1	1	1	1	1					
m9										1		1	1	1	1	1	1	1	
m2											1	1					1		
m6				1		1					1			1	1	1	1	1	1
m4										1	1	1	1			1	1	1	1
m8										1	1	1			1	1	1	1	1

values, called *levels*, for the parameters, called *factors*, of the decision problem.

Table 9 reports the adopted levels for each factor in the experimental analysis.

Figures 8 to 10 present the *main effects plot* (Minitab ® Statistical Software Inc.) for the following performance indices: η_G, called η(G) in figures, τ, BE.

Similarity indices perform in a different way in terms of η_G, τ, BE. In particular problem oriented (PO) perform better than general purpose (GP). Clink rule and percentile threshold value equal to 50° (or 75°) seem to be the best levels to set the clustering algorithm. The best performing indices are Seiffoddini - S (1987) and Nair and Narendran - N (1998).

Figure 8. Main effects plot for grouping measure

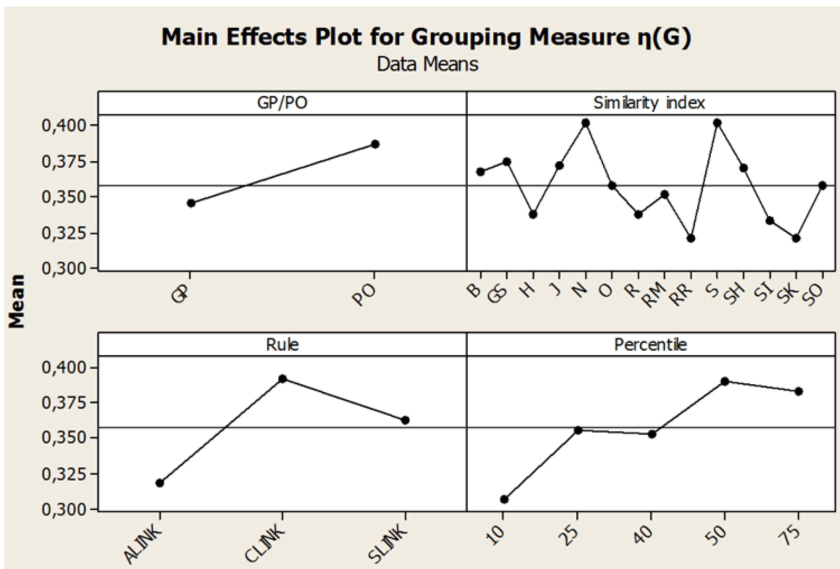


Figure 9. Main effects plot for grouping efficacy

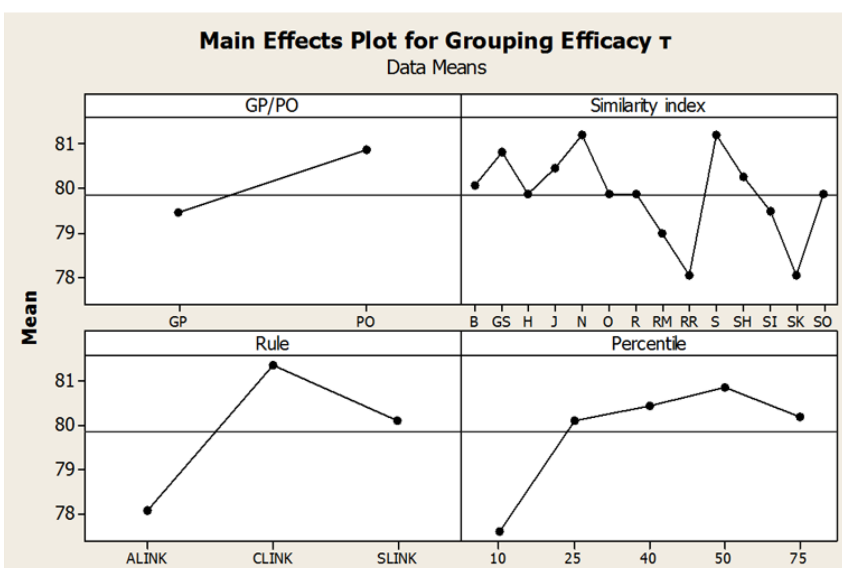


Figure 10. Main effects plot for bond efficiency

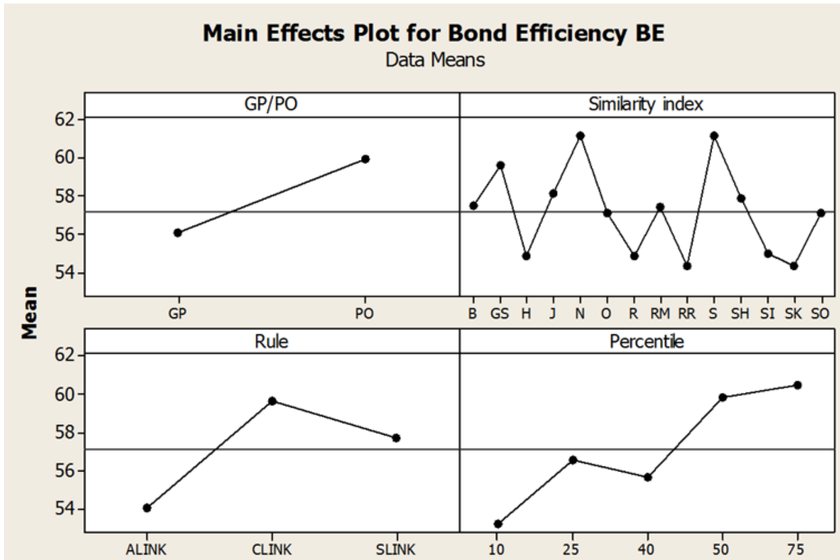


Figure 11 shows that the number of exceptional elements significantly depends on the adopted threshold value of group similarity, but the adopted similarity index is not important.

η_{QI}, called η(QI) in Figure 12, has an anomalous trend if compared with previous graphs.

Figure 13 shows the trend of the EE for different values of couples of factors, and the importance of the percentile threshold value of group similarity.

Similarly, Figure 14 shows the importance of threshold value of similarity and clink rule for grouping items.

Figure 11. Main effects plot for exceptional elements

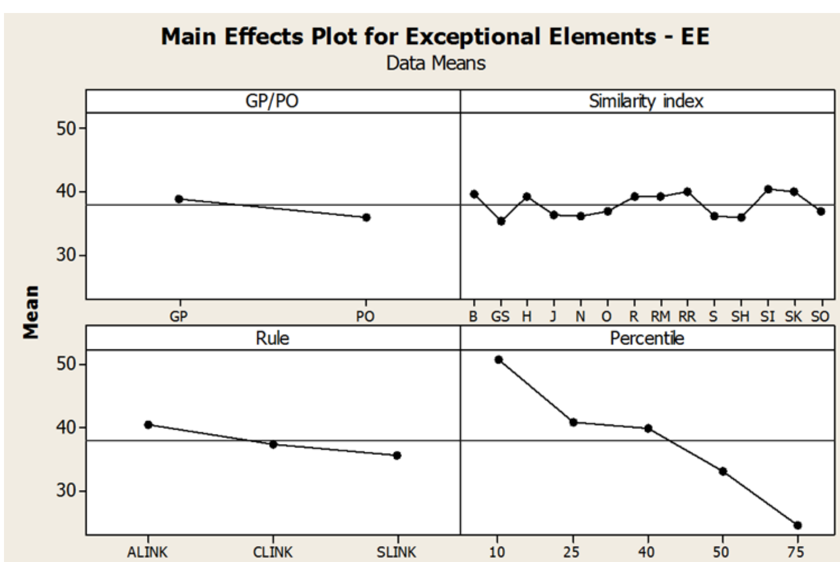
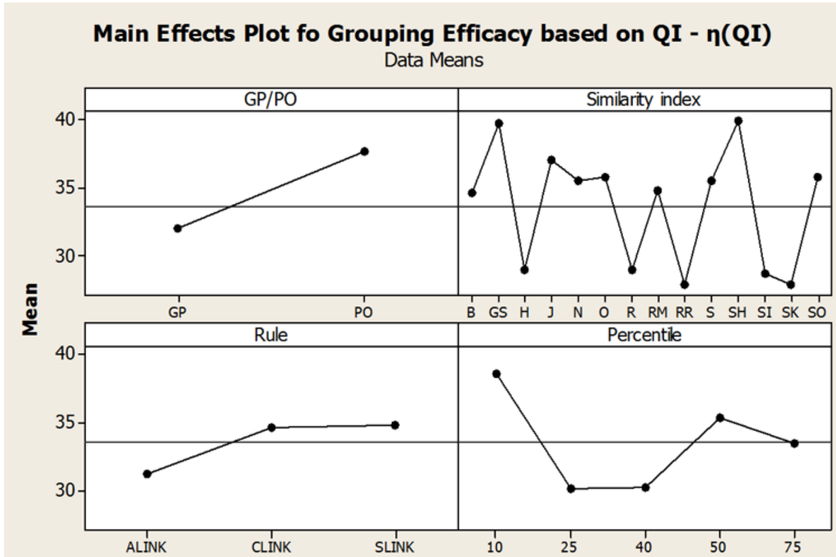


Figure 12. Main effects plot for grouping efficacy based on QI



CONCLUSION AND FURTHER RESEARCH

This chapter illustrates the CFP as supported by the similarity based manufacturing clustering, and a hierarchical and systematic procedure for supporting managers in the configuration of cellular manufacturing systems by the application of cluster analysis and similarity indices. In particular, both general purpose and problem oriented indices are illustrated and applied. The experimental analysis conducted on a literature problem oriented case study represents the first basis for the identification of the best setting of the cell formation problem and supporting decision models and tools.

Figure 13. Exceptional elements for couples of factors

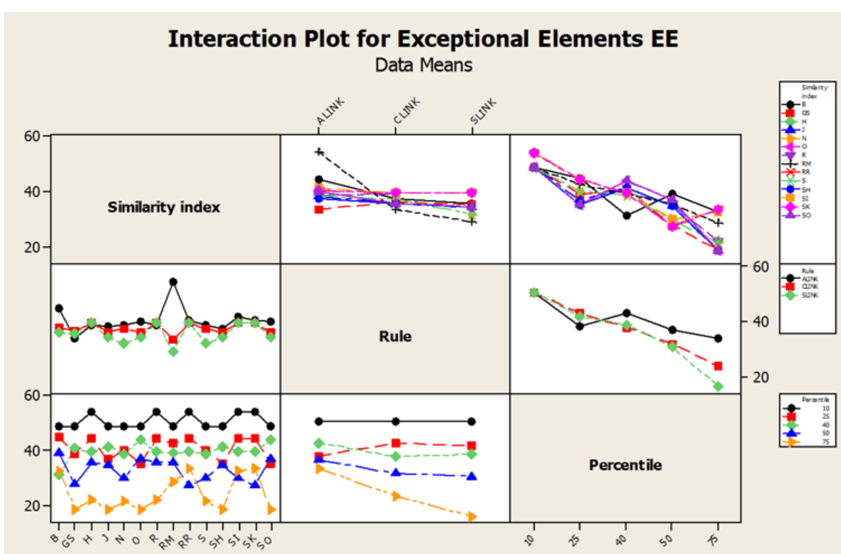
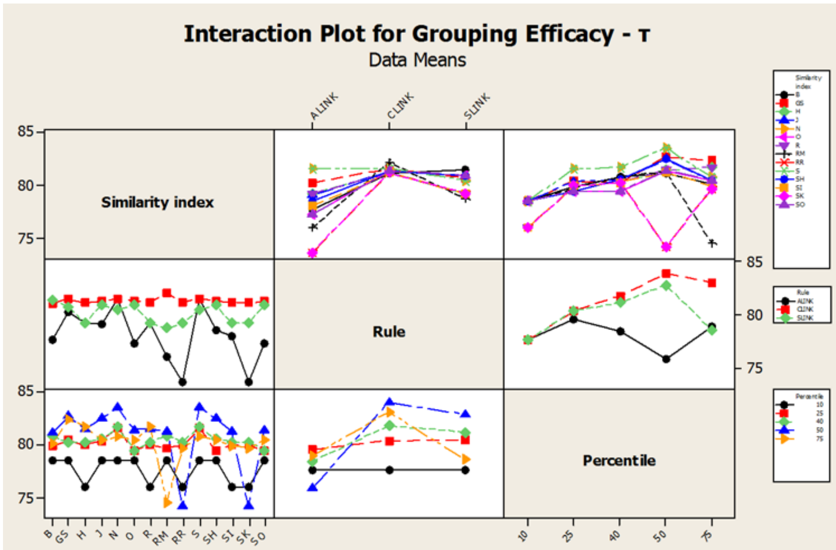


Figure 14. Interaction plot for τ



For the first time, this chapter successfully applies the threshold group similarity index to problem oriented similarity environment. The threshold value was introduced by the authors in a previous study on general purpose indices evaluation (Manzini et al. 2010).

This chapter confirms the importance of this threshold cut value for the dendrogram when it is explained in percentile on the number of nodes.

Further research is expected to improve the experimental analysis including more case studies and applications. Finally it is important to improve the critical process of part family formation and the decisions regarding the duplication of machines and resources in different manufacturing cells in order to minimize intercellular flows.

REFERENCES

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. (Sage University Paper series on Quantitative Applications in the Social Sciences, 1984, No. 07-044), Beverly Hills, CA: Sage.

Alhourani, F., & Seifoddini, H. (2007). Machine cell formation for production management in cellular manufacturing systems. *International Journal of Production Research*, *45*(4), 913–934. doi:10.1080/00207540600664144

Bindi, F., Manzini, R., Pareschi, A., & Regattieri, A. (2009). Similarity-based storage allocation rules in an order picking system. An application to the food service industry. *International Journal of Logistics Research and Applications*, *12*(4), 233–247. doi:10.1080/13675560903075943

De Witte, J. (1980). The use of similarity coefficients in production flow analysis. *International Journal of Production Research*, *18*, 503–514. doi:10.1080/00207548008919686

Gupta, T., & Seifoddini, H. (1990). Production data based similarity coefficient for machine-component grouping decisions in the design of a cellular manufacturing system. *International Journal of Production Research*, *28*, 1247–1269. doi:10.1080/00207549008942791

Heragu, S. (1997). *Facilities design*. Boston, MA: PWS Publishing Company.

Kumar, C. S., & Chandrasekharan, M. P. (1990). Grouping efficacy a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology. *International Journal of Production Research*, *28*(2), 233–243. doi:10.1080/00207549008942706

Manzini, R., & Bindi, F. (2009). Strategic design and operational management optimization of a multi stage physical distribution system. *Transportation Research Part E, Logistics and Transportation Review*, *45*, 915–936. doi:10.1016/j.tre.2009.04.011

Manzini, R., Bindi, F., & Pareschi, A. (2010). The threshold value of group similarity in the formation of cellular manufacturing system. *International Journal of Production Research*, *48*(10), 3029–3060. doi:10.1080/00207540802644860

Manzini, R., Persona, A., & Regattieri, A. (2006). Framework for designing and controlling a multicellular flexible manufacturing system. *International Journal of Services and Operations Management*, *2*, 1–21. doi:10.1504/IJSOM.2006.009031

McAuley, J. (1972). Machine grouping for efficient production. *Production Engineering*, *51*, 53–57. doi:10.1049/tpe.1972.0006

Mosier, C. T. (1989). An experiment investigating the application of clustering procedures and similarity coefficients to the GT machine cell formation problem. *International Journal of Production Research*, *27*(10), 1811–1835. doi:10.1080/00207548908942656

Nair, G. J., & Narendran, T. T. (1998). CASE: A clustering algorithm for cell formation with sequence data. *International Journal of Production Research*, *36*, 157–179. doi:10.1080/002075498193985

Papaioannou, G., & Wilson, J. M. (2010). The evolution of cell formation problem methodologies based on recent studies (1987-2008): Review and directions for future research, (vol. 206, pp. 509-521).

Sarker, B. R. (2001). Measures of grouping efficiency in cellular manufacturing systems. *European Journal of Operational Research*, *130*, 588–611. doi:10.1016/S0377-2217(99)00419-1

Seifoddini, H. (1987). Incorporation of the production volume in machine cell formation in group technology applications. *Proceedings of the 9th International Conference on Production Research ICPR*, (pp. 2348-2356).

Seifoddini, H., & Djassemi, M. (1994). Analysis of efficiency measures for block diagonal machine-component charts. *Proceedings of the 16th International Conference on Computers and Industrial Engineering*, Ashikaga, Japan.

Seifoddini, H., & Djassemi, M. (1996). The thresold value of a quality index for formation of cellular manufacturing systems. *International Journal of Production Research*, *34*(12), 3401–3416. doi:10.1080/00207549608905097

Sokal, R. R., & Sneath, P. H. A. (1968). *Principles of numerical taxonomy*. San Francisco, CA: W. H. Freeman.

Stawowy, A. (2004). Evolutionary strategy for manufacturing cell design. *Omega: The International Journal of Management Science*, *34*, 1–18. doi:10.1016/j.omega.2004.07.016

Yin, Y., & Yasuda, K. (2006). Similarity coefficient methods applied to cell formation problem: A taxonomy and review. *International Journal of Production Economics*, *101*, 329–352. doi:10.1016/j.ijpe.2005.01.014

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 140-163, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 30
Performance Comparison of Cellular Manufacturing Configurations in Different Demand Profiles

Paolo Renna
University of Basilicata, Italy

Michele Ambrico
University of Basilicata, Italy

ABSTRACT

Cellular manufacturing systems (CMSs) are an effective response in the economic environment characterized by high variability of market. The aim of this chapter is to compare different configurations of cellular models through the main performance. These configurations are fractal CMS (defined FCMS) and cellular systems with remainder cells (defined RCMS), compared to classical CMS used as a benchmark. FCMSs consist of a cellular system characterized by identical cells each capable of producing all types of parts. RCMSs consist of a classical CMS with an additional cell (remainder cell) that in specific conditions is able to perform all the technological operations. A simulation environment based on Rockwell ARENA® has been developed to compare different configurations assuming a constant mix of demand and different congestion levels. The simulation results show that RCMSs can be a competitive alternative to traditional cells developing opportune methodologies to control the loading of the cells.

INTRODUCTION

Competitiveness in today's market is much more intense compared to the past decades. Considerable resources are invested on facilities planning and re-planning in order to adapt the manufacturing systems to the market changes. A well-established manufacturing philosophy is the group technology concept.

Group technology (GT) can be defined as a manufacturing philosophy identifying similar parts and grouping them together to take advantage

DOI: 10.4018/978-1-4666-1945-6.ch030

of their similarities in manufacturing and design (Selim et al.,1998). It is the basis of so-called cellular manufacturing systems (CMSs). In current production scenario demand for products is characterized by continuous fluctuations in terms of volumes, type of product (part mix), new products introduction and the life cycle of products has significantly reduced. The planning horizon needs to be divided into smaller horizons (time bucket) and the length of each period is related to the characteristics of products. These characteristics need to be considered in design process of a manufacturing system. Introduction of Cellular Manufacturing Systems has already introduced significant improvements. They are conceived with the aim of reducing costs such as setup costs or handling costs and also to reduce lead time and work in process (WIP). They combine advantages of flow shop and job shop, but a further step can be accomplished to be competitive in the market. They allow significant improvements such as: product quality, worker satisfaction, space utilization. Benefits and disadvantages (Irani et al.,1999) are showed in Table 1. They documented that companies implementing cellular manufacturing have a very high probability of obtaining improvements in various areas.

The first column of Table 1 shows the case studies with improvements and the second column reports the percentage of improvement of the measures. Similarly, the third column shows the percentage of cases with worsening and in the fourth column is evidenced the rate of deterioration.

The demand volatility and continuous new product introduction lead to re-configure several times the cellular manufacturing systems in order to keep a high level of performance.

For the above reasons, new configurations have been proposed in literature such as Virtual Cell Manufacturing System (VCMS), Fractal Cell Manufacturing System (FCMS), Dynamic Cell Manufacturing System (DCMS), with the aim of keeping high flexibility of manufacturing systems.

The concept of DCMS was introduced for the first time by Rehault et al. (1995). It provides a physical reconfiguration of the cells. The reconfiguration activity can be periodic or resulting from the variation of performance parameters. Reconfigure can mean duplicating machines,

Table 1. Benefits and disadvantages of CMS

Measure	Percentage cases with improvements	Average percentage improvement	Percentage cases with worsening	Average percentage worsening
Tooling cost	31%	-10%	69%	+17%
Labor cost	91%	-33%	9%	+25%
Setup Time	84%	-53%	16%	+32%
Cycle Time	84%	-40%	16%	+30%
Machine utilization	53%	+33%	47%	-20%
Subcontracting	57%	-50%	43%	+10%
Product quality	90%	+31%	10%	-15%
Worker satisfaction	95%	+36%	5%	-
Space utilization	17%	-25%	83%	+40%
WIP inventory	87%	-58%	13%	+20%
Labor turnover/absenteeism	100%	-50%	0	-
Variable production cost	93%	-18%	7%	+10%

relocating machines between cells, removing machines, or also subcontracting some parts to other companies. These problems must be addressed by the decision maker.

The concept of VCMS requires that the machines are dedicated to a part family but these machines are not necessarily close together in a classical cell. One machine can belong simultaneously to different cells. Hence sharing of machine makes the system more flexible. Moreover the machines are not shifted as dynamic cellular system therefore costs of reallocation are eliminated. On the other hand we must consider the increase in the movements of parts (or batches) across machines. A further problem may be the complication in the measurement of performance of the cells. This is because monitoring stations are usually located out of the cell, but in this case the cell does not exist physically.

The FCMSs are based on the constructions of identical cells and they are not built for different families. The idea comes from Skinner (1974) and that is to build factory within a factory with duplication of processes. Each cell can work all products. Working time will be greater but these configurations are very effective if there are changes in part mix and in cases of machine breakdowns. Even if for example there are flash orders.

A further idea was mentioned by Sripathy Maddisetty (2005). The author referred to so-called remainder cells and we can call them RCMSs. In addition to traditional cells refer to the product families you may create an additional cell that operates when conditions such as machine failures or overloaded machines occur. Focusing on an advanced design the RCMSs could provide interesting results in terms of competitiveness.

Our goal in this chapter is to compare the various approaches to the design of manufacturing systems, making a complete performance comparison. In particular we aimed to compare the following systems: CMSs, FCMSs and RCMSs. A simulation environment has been developed to compare the performance (WIP, Throughput Time, Tardiness, Throughput and Average Utilization) using as a benchmark the classic CMS. The aim is to evaluate the responses of different systems when market fluctuations occur in terms of arrival demand. The chapter is structured as follows. Section 2 provides an overview of the literature of various manufacturing system configurations, while in section 3 the system context is formulated. In section 4 there is a brief description of scheduling approaches. Section 5 presents the simulation environment and the case study while in section 6 are discussed simulation results. In Section 7 conclusions and future developments are discussed.

BACKGROUND

Recently, several authors have investigated the configuration of manufacturing cells in order to keep a high level of performance when the market conditions change.

Hachicha et al. (2007) proposed a simulation based methodology which takes into consideration the stochastic aspect in the CMS. They took into account the existence of exceptional elements between the parts and the effect of the correspondent inter-cell movements. They compared two strategies: permitting intercellular transfer and exceptional machine duplication. They used the simulation (Rockwell Arena) and they analyzed the following performance: mean transfer time, mean machining time, mean wait time, mean flow time. They assumed demand fixed and known for the parts. They did not consider failures of machines and maintenance policies.

A multi-objective dynamic cell formation was presented by Bajestani et al. (2007) where purpose was to minimize simultaneously total cell load variation and sum of miscellaneous costs (machine cost, inter-cell material handling cost, and machine relocation cost). Since the problem is *NP-hard* they used a scatter search approach for finding locally Pareto-optimal frontier.

Safei et al. (2007) proposed to use an approach based on fuzzy logic for the design of CMS under uncertain and dynamic conditions. They began by finding that in most of research related on DCMS input parameters were considered deterministic and certain. Therefore they introduced fuzzy logic as a tool for the expression of the uncertainty in design parameters such as part demand and available machine capacity.

Ahkioon et al. (2007) tried to investigate DCMS focusing on routing flexibility. They studied the creation of alternate contingency process routings in addition to alternate main process routings for all part types. Contingency routings had the function to provide continuity in case of exceptional events such as machine breakdowns but also flash orders. Furthermore their work provided discussions on the trade-off between the additional cost related to the formation of contingency routings and the advantages of increased flexibility. Linearized model proposed by the authors was solved with CPLEX.

Aryanezhad et al. (2008) developed a new model which simultaneously embrace dynamic cell formation and worker assignment problem. They focused on two separate components of cost: the machine based costs such as production costs, inter-cell material handling costs, machine costs and human related costs such as hiring costs, firing costs, training costs and wages. They made the comparison of two models. One considered the machine costs and the other considered both machine costs and human related costs. The model was *NP-hard* even though they did not consider learning curve.

Xiaoqing Wang et al. (2008) proposed a nonlinear multi-objective mathematical model in dynamic cells formation problem by giving weighing to three conflicting objectives: machine relocation costs, utilization rate of machine capacity, and total number of inter-cell moves over the planning horizon. A scatter search approach was developed to solve the nonlinear model. Results were compared with those obtained by CPLEX. They considered certain demand and they did not consider machine breakdowns.

Safei et al. (2009) proposed an integrated mathematical model of the multi-period cell formation and production planning in a dynamic cellular manufacturing system (DCMS). The focus was on the effect of the trade-off between production and outsourcing costs on the reconfiguration of the cells.

Balakrishnan (2005) discussed cellular manufacturing system under conditions of changing product demand. He made a conceptual comparison to virtual cell manufacturing and he discussed a case study.

Kesen et al. (2008) investigated three different types of system (cellular layout, process layout and virtual cells) by using simulation. They paid attention to the following performance: mean flow time and mean tardiness. Based on these simulations they used regression meta-models to estimate the systems behaviours. They only considered one family-based scheduling scheme and they did not consider extraordinary events such as machine failures.

Vakharia et al. (1999) proposed and validated analytical approximations for comparing the performance of virtual cells and multistage flow shops. First they used these approximations and hypothetical data to identify some key factors that influenced the implementation of virtual cells in a multistage flow shop environment. Then they concluded with an application of approximations to industrial data.

Kesen et al. (2009) examined the behaviours of VCMs, process layouts and cellular layouts. They addressed the VCMs by using family-based scheduling rule. The different systems were compared by simulation. Subsequently they developed an ant colony optimization based meta-models to reflect the system's behaviours.

Kesen et al. (2010) presented a genetic algorithm based heuristic approach for job scheduling in virtual manufacturing cells (VMCs). Cell configurations were made to optimize the scheduling objective under changing demand conditions. They considered the case with multiple jobs and different processing routes. It was considered multiple machine types with several identical machines in each type and they were located in different locations in the shop floor. The objective was to minimize the total travelling distance. To evaluate the effectiveness of the genetic algorithm heuristic they compared it with a mixed integer programming solution. Results showed that genetic algorithm was promising in finding good solutions in very shorter.

Uday Venkatadri et al.(1997) proposed a methodology for designing job shops under the fractal layout organization as an alternative to the more traditional function and product organizations. The challenge in assigning flow to workstation replicates was that flow assignment is in itself a layout dependent decision problem. They proposed an iterative algorithm that updated layouts depending on flow assignments, and flow assignments based on layouts. Their work has had the far-reaching consequence of demonstrating the validity of the fractal layout organization in manufacturing systems (FCMSs).

Montreuil (1999) developed a new fractal alternative for manufacturing job shops which allocated the total number of workstations for most processes equally across several fractal cells. He introduced fractal organization and he briefly discussed the process of implementing fractal designs. He illustrated a case example and he showed that system is characterized by great flexibility.

Maddisetty (2005) discussed the design cells in a probabilistic demand environment. He discussed idea of remainder cells (RCMS). A remainder cell is a kind of lung to cope in changes in demand. He examined the following performance: total WIP, average flow time, machine utilization. He proposed a comparison using three different approaches: mathematical, heuristic, and simulation.

Süer et al. (2010) proposed a new layered cellular manufacturing system to form dedicated, shared and remainder cells to deal with the probabilistic demand. Moreover they proposed a comparison of its performance with the classical cellular manufacturing system. Simulation and statistical analysis were performed to help identify the best design within and among both layered cellular design and classical cellular design. They observed that the average flow time and total WIP were not always the lowest when additional machines were used by the system, but the layered cellular system performed better when demand fluctuations was observed.

There are several limitations encountered in existing literature. In previous research the demand of products was usually determined at the beginning of each period and it was known. The change in part mix was rarely assumed. Frequently the bottleneck station in each cell was considered as fixed and independent of the type of the part. Almost never were held in account exceptional events such as machine failures and maintenance. Almost never flash orders was considered and similarly backorders. The concept of learning curve was rarely covered. Furthermore hardly researchers focused on a wide range of performance measures.

In this chapter the objective is to evaluate the reaction of different manufacturing systems configurations (CMSs, FCMSs and RCMSs) when there is a fluctuation in terms of arrival demand. The configurations are investigated considering the same machines for all cases; the machines are set in order to obtain the particular configuration. The analysis conducted allows to highlight the most promising configurations in terms of performance measures. Another objective of the chapter is to develop a simulation environment based on Rockwell Arena® tool in order to analyse the different configurations. The simulation allows build a model with minor simplification compared to mathematical models which require significant simplifications (linearization) in cases of complex systems. Moreover the dynamic model (demand not known a priori, unexpected events like machine breakdowns) cannot be obtained with mathematical models.

MANUFACTURING SYSTEM CONTEXT

The mentioned objective of this chapter is to compare the performance of different manufacturing systems. In particular, the configurations analyzed by using simulation tools based on the software Rockwell ARENA® are: CMS, FCMS and RCMS. Moreover another configuration has been considered changing the layout of machines and obtaining a CMS in line.

The manufacturing system consists of *M* machines general purpose that are used for each configuration. It has been considered three part families. We consider a constant mix of each part family.

We introduce the following assumptions for the model:

- the demand for each part type is unknown at priori end it is extracted randomly from an exponential distribution. Therefore, the parameter to set is the exponential parameter;
- set-up times are not simulated. When, the manufacturing cells are configured the set-up times are very low for the product family assigned to the cell;
- the due date is obtained by processing time multiplied with an index greater than or equal to 1;
- Machine breakdowns and maintenance are not considered;
- intra-cell handling times are negligible;
- it is assumed that parts moved in units;
- each configuration presents the same number of machines in order to make a comparison in the same conditions.

The performance measures used to compare the manufacturing systems are the following:

- Work in Process (WIP);
- Average utilization of the manufacturing system;
- Throughput time;
- Average throughput time;
- Tardiness (total of all the parts);
- Throughput.

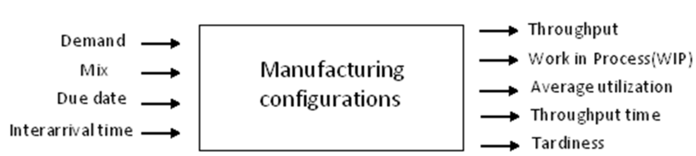
Figure 1 describes the parameters and the performance analyzed in this research.

The first manufacturing system configuration considered is a classical cellular system (CMS). The scheme is showed in Figure 2.

The system manufactures *N* product families with *N* cells. Each cell is specialized to perform the technological operations required by the product family assigned (setup time is not necessary). In this chapter, it has been also considered a CMS with a different routing, as showed in Figure 3.

The second configuration considered is the FCMS. In this case, the allocation of machines to cells is performed in order to obtain N identical cells. Each cell manufactures all product families with higher processing time, because the machines will be able to perform all the technological op-

Figure 1. Manufacturing configurations analysis



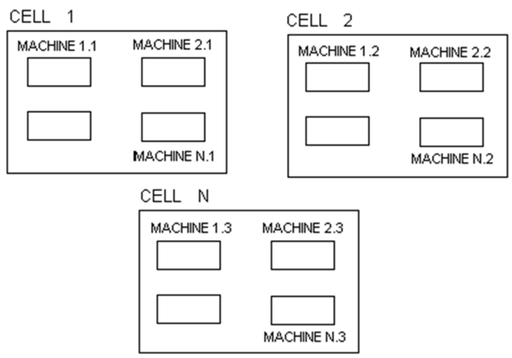


Figure 2. CMS configuration

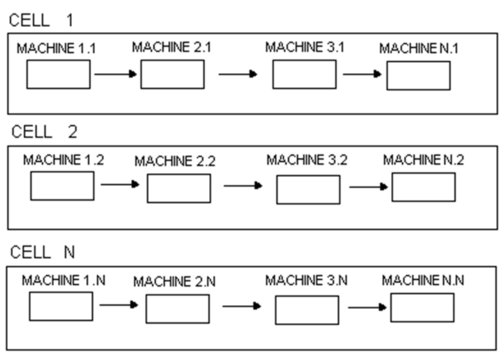


Figure 3. CMS in line configuration

erations required. The scheme of FCMS is showed in Figure 4.

The third configuration considered is the RCMS. In this configuration there are *N* cells respectively for *N* product families. In addition, there is a further cell called remainder cell where all operations can be performed with higher processing times. It may be useful in case of machine failures but also in case of congestion of the system. The scheme of RCMS is showed in Figure 5.

Each configuration includes the same number of machines and the time to manufacture each part is assumed the same, except for fractal cells (belonging to FCMS) and the remainder cell (belonging to RCMS) where machines can produce all kinds of part with a higher processing time (general purpose machine configuration). Therefore the processing time of machine *i-th* in fractal cell (pt_{if}) and processing time of machine *i-th* in remainder cell (pt_{ir}) is major of processing time of the machine *i-th* in the cell *j-th* in CMS (machine configured for the technological operations of a particular family) (pt_{ij}):

$$over = \frac{pt_{if}}{pt_{ij}} = \frac{pt_{ir}}{pt_{ij}}, over > 1 \qquad (1)$$

LOADING POLICY

In the previous section we have discussed the different cell configurations. Each configuration needs a loading approach policy to operate.

For classical CMS parts arrive in the system and each family has its own cell competence. In CMS we have provided two different layouts: one

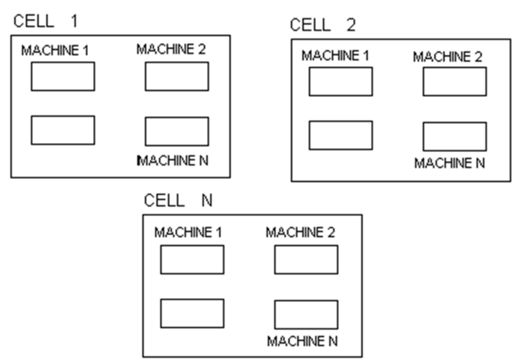


Figure 4. FCMS configuration

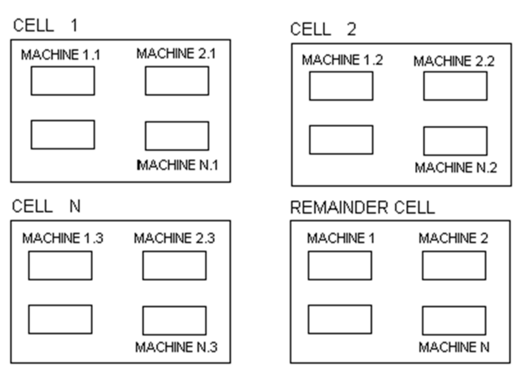


Figure 5. RCMS configuration

with parallel machines and other with machines in line, as described above.

In FCMS configuration parts arrive in the system and they are routed to cells with minor workload.

The RCMS needs a specific loading policy for the use of the remainder cell. Parts arrive in the system and each cell is designed for a part family. In each cell, there is a controller that adopts the following strategy: it measures the number of parts in queue in each machine. If the measured value in cell *j-th* is greater than a maximum threshold of the cell (defined $Smax_j$) then the part is conveyed to the remainder cell. Similarly, when the measured value is minor of a minimum threshold of the cell *j-th* (defined $Smin_j$) then the part is assigned to the cell designed for the part family. The logic of controller above described is showed in the flowchart of Figure 6.

SIMULATION ENVIRONMENT

The manufacturing system consists of *M=10* machines. All different configurations are obtained re-allocating the same number of machines available. It is considered that each machine functions for 24 hours a day. Therefore total numbers of minute that system works is considered to be 43200 minutes per month. This is the simulation horizon considered.

In order to evidence only the difference among the configurations, it is assumed that each part needs 40 minutes to complete processing. This technological time is divided by the number of machines used in the process, depending on manufacturing configuration. As above introduced it is equal for all parts except for those made in fractal cells and remainder cell where machines take more time.

We assume three product families. The product mix is as follow: Product 1 (40%), Product 2 (40%) and Product 3 (20%).

We have analyzed the performance of four different cellular systems changing one parameter: the average inter-arrival time. We have considered five different values of inter-arrival time that leads to different congestion levels of the manufacturing system (see Table 2).

These values were selected to keep the average utilization of machines in a range that goes from 0.56 (low utilization) to 0.99 (high utilization).

The demand for each part type is unknown at priori and it is extracted randomly from an

Figure 6. The logic of RCMS

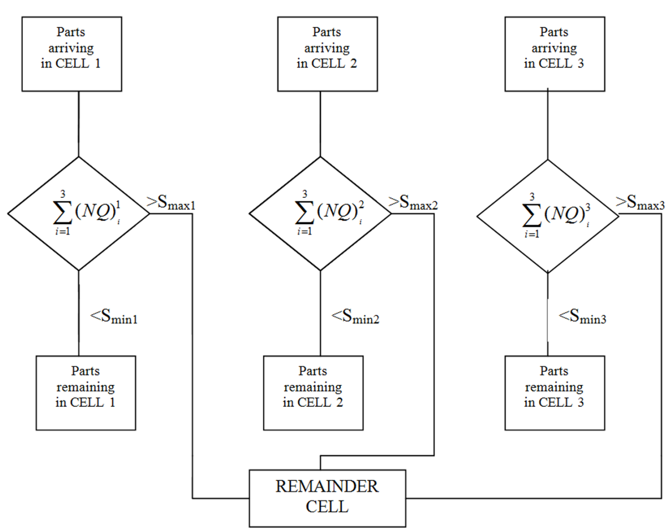


Table 2. Average inter-arrival times

4
4.5
5
6
7

exponential distribution with mean equal to the inter-arrival time reported in Table 2.

The due date is obtained from the sum of the arrival time (*tnow*) and the technological working time (*WT*) multiplied with an index (DdateINDEX), as showed in equation 2.

$$Ddate = tnow + (WT \cdot Ddate_{INDEX}) \qquad (2)$$

The *WT* is obviously equal to 40 minutes. The $Ddate_{index}$ is 1.5 for parts 1 and 2, while it is 1 for part 3. The minor index of part 3 is justified by the lower demand than other part-mix, so there is no shift of the due date. However the due dates are the same for all configurations examined. Therefore not affect the comparison, but they are included in the model for completeness.

Cellular systems analyzed are those already mentioned: CMS, CMS in line, RCMS and FCMS. The benchmark system is the CMS. The simulation environment has been developed by Rockwell Arena® tool.

Arena is characterized by a block diagram that makes it more familiar environment simulation. The arrival stations of the parts and the exit station are showed in the Figure 7.

In the first three boxes are showed the arrival stations where to each part is assigned a delivery time and a destination in the respective cell for processing; then the parts leave the arrival station.

Exit station is equal for all types of configuration: if the delivery time has been observed then the WIP is updated and the part leaves the system. Otherwise the delay is calculated.

Cellular Manufacturing System

In this case we consider three cells of production. The first two cells containing four identical machines working in pairs and in parallel. These cells are respectively for both products type 1 and type 2. The third cell contains 2 machines for products of type 3 (minor product mix). Each machine has a process time equal to 20 minutes. The scheme is showed in Figure 8.

In each rectangle is indicated the working time.

Cellular Manufacturing System in Line

In this case we also consider 3 cells of production. The first two cells containing 4 machines in line. Each machine has a process time equal to 10 minutes. These cells are respectively for type 1 and type 2. The third cell contains 2 machines for product type 3, each machine has a process

Figure 7. Arrival and exit stations

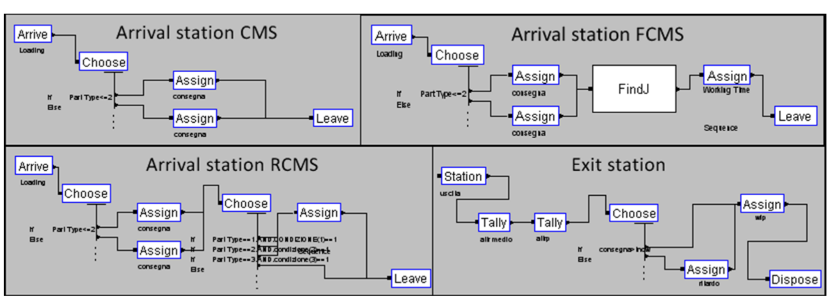


Figure 8. CMS considered in simulation

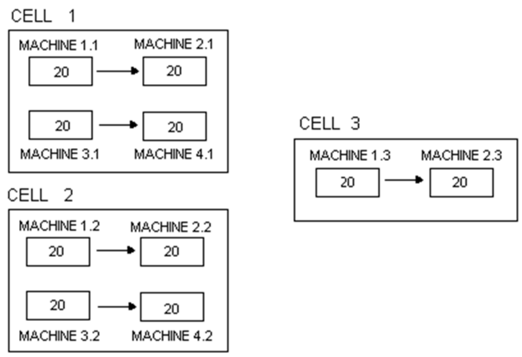
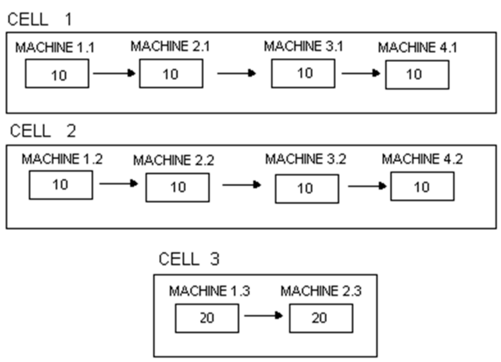


Figure 9. CMS in line considered in simulation



time equal to 20 minutes. The scheme is showed in Figure 9.

Fractal Cellular Manufacturing System

In this case there are 5 identical cells. Each cell contains 2 machines and each cell is able to work on all the product mix. The scheme is showed in Figure 10.

Naturally the machines perform the manufacturing operations with a major process time (see equations 1 and 2) because they are not dedicated to a part family but they are configured for all operations. In fact the process time of each machine is equal to 20 units time increased by 20% (over=1.2).

Remainder Cellular Manufacturing System

In this case there are 3 cells (one for each part type) and there is a remainder cell where is defined a loading policy based on the number of parts in queue in other cells. The scheme is showed in Figure 11.

The three machines operating in cell 1 (product type 1) has a process time equal to 13,33 minutes. The same for the machines operating in cell 2 (product type 2). The two machines operating in cell 3 has a process time equal to 20 minutes. The machines assigned to the remainder cell perform the manufacturing operations with a

Figure 10. FCMS considered in simulation

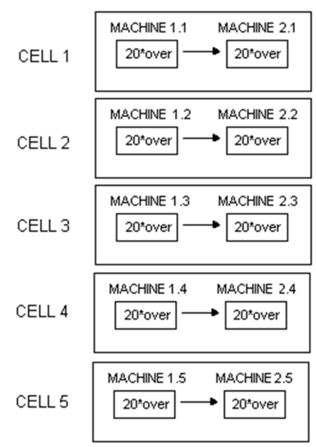
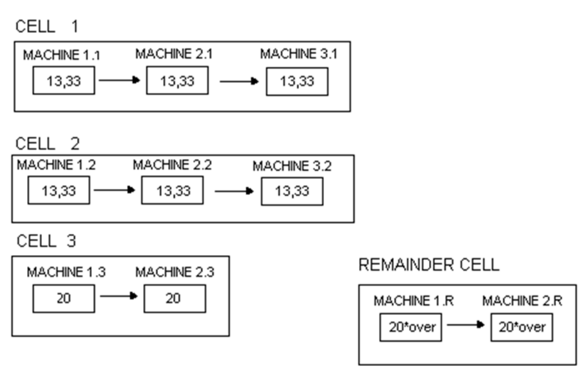


Figure 11. RCMS considered in simulation



major process time (see equations 1 and 2) because they are configured for all operations; the process time of each machine is equal to 20 units time increased by 20%(over=1.2).

In this work, it has been investigated different instances of the same policy loading about the use of remainder cell. Each cell has a controller that measures the number of parts in queue in each machine. Using thresholds the parts can be conveyed to the remainder cell. In ARENA the controller is showed in Figure 12.

The first *"scan"* controls the maximum threshold ($Smax_j$) and therefore assigns the part to the cell. Similarly, the second "*scan"* checks the minimum thresholds($Smin_j$).

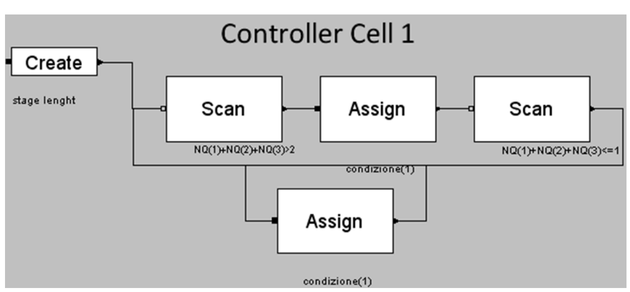
For the values of maximum ($Smax_j$) and minimum ($Smin_j$) thresholds have been considered respectively six cases, equal for all three cells (see Table 3).

SIMULATION RESULTS

The length of each simulation is fixed to 43200 minutes. During this period the average inter-arrival time and part mix are both constant. Table 4 reports the design of simulation experiments conducted for all four configurations of the manufacturing system.

Combining the five inter-arrival times, four system configurations, and for the last configuration (RCMS) six cases regarding the thresholds, it has been obtained 45 experimental classes.

Figure 12. Control blocks cell 1



For each experiment class have been conducted a number of replications able to assure a 5% confidence interval and 95% of confidence level for each performance measure.

As previously described the performance measures investigated are the following:

- Work in Process (WIP);
- Average utilization of the manufacturing system (*av.utilization*);
- Throughput time for each part j(*thr. Time j*);
- Average throughput time (*average thr. Time*);
- Total tardiness time of all the parts (*tardiness*);
- Throughput (*thr.*).

The objective of the analysis of simulation results is the comparison between different manufacturing configurations and classical cellular configuration (CMS, used as base for percentage

Table 3. Threshold values

Cases	Smax	Smin
1	7	5
2	5	3
3	3	2
4	4	1
5	3	1
6	2	1

Table 4. Experimental classes

Exp. No.	Configuration	Inter-arrival	Exp. No.	Configuration	Inter-arrival
1	CMS	4	26	RCMS(3,2)	4
2	CMS	4,5	27	RCMS(3,2)	4,5
3	CMS	5	28	RCMS(3,2)	5
4	CMS	6	29	RCMS(3,2)	6
5	CMS	7	30	RCMS(3,2)	7
6	CMS in line	4	31	RCMS(4,1)	4
7	CMS in line	4,5	32	RCMS(4,1)	4,5
8	CMS in line	5	33	RCMS(4,1)	5
9	CMS in line	6	34	RCMS(4,1)	6
10	CMS in line	7	35	RCMS(4,1)	7
11	FCMS	4	36	RCMS(3,1)	4
12	FCMS	4,5	37	RCMS(3,1)	4,5
13	FCMS	5	38	RCMS(3,1)	5
14	FCMS	6	39	RCMS(3,1)	6
15	FCMS	7	40	RCMS(3,1)	7
16	RCMS(7,5)	4	41	RCMS(2,1)	4
17	RCMS(7,5)	4,5	42	RCMS(2,1)	4,5
18	RCMS(7,5)	5	43	RCMS(2,1)	5
19	RCMS(7,5)	6	44	RCMS(2,1)	6
20	RCMS(7,5)	7	45	RCMS(2,1)	7
21	RCMS(5,3)	4			
22	RCMS(5,3)	4,5			
23	RCMS(5,3)	5			
24	RCMS(5,3)	6			
25	RCMS(5,3)	7			

computation). The aim is to use the performance parameters to highlight the behaviour of different configurations when changing the volume of demand (the variation of average inter-arrival times).

Table 5 shows the average utilizations of machines in classical CMS at different inter-arrival times.

Therefore the simulations are performed for five congestion levels of the manufacturing system. It is important to emphasize that the results showed do not include machine breakdowns.

Table 6 reports the first three parameters (WIP, Tardiness and Throughput) for the different manufacturing configurations. Table 6 shows the average values over inter-arrival times with the respective standard deviations (St.dev). The standard deviation allows to highlight the variability of the results when the inter-arrival changes. The

Table 5. Average utilizations

Configuration	Inter-arrival time	Av. utilization
CMS	4	0,99
	4,5	0,88
	5	0,80
	6	0,66
	7	0,57

percentages refer to the comparison with the classical CMS. The positive percentages represent an increase of the respective factor while the negative percentages represent a decrease. Table 7 is the same for the throughput time of different parts and for the average throughout time.

Tables 6 and 7 show that CMS with configuration in line has almost the same behaviour of the classical CMS except for the tardiness that increments significantly.

Tables 6 and 7 also show that fractal configuration (FCMS) is the worst configuration. This is because the scheduling policy used is more simply. An opportune policy needs to be implemented for the FCMS. This is a limit of FMCS configuration, because a more complex control system has to be designed. The standard deviation shows the variability of the performance measures related to the inter-arrival changes in fact the FCMS is the configuration with the higher dependence on the inter-arrival changes. As the reader can notice, the RCMS performance depends on the choose of the threshold values.

Table 8 reports the variation of performance observed in correspondence of three values of inter-arrival times (5, 6, and 7). The percentages always refer to the comparison with the classical CMS.

Among the various configurations of RCMS is showed only one (with thresholds 2, 1) with the most interesting results (see Table 8). Except for value of tardiness (when inter-arrival time is

Table 6. Simulation results

	WIP		Tardiness		Throughput	
	average	St. dev	average	St. dev	average	St. dev
CMS(in line)	2,15%	1,62%	85,97%	179,28%	0,01%	0,19%
FCMS	495,96%	699,08%	956,98%	1583,56%	-4,49%	6,74%
RCMS 7,5	62,55%	64,65%	148,50%	49,65%	-0,77%	1,66%
RCMS 5,3	76,76%	91,60%	136,93%	49,38%	-0,99%	2,10%
RCMS 3,2	107,54%	118,07%	134,58%	71,72%	18,64%	45,45%
RCMS 4,1	95,70%	133,95%	133,78%	96,78%	-1,47%	2,92%
RCMS 3,1	132,86%	170,05%	191,64%	237,27%	-1,62%	3,36%
RCMS 2,1	203,37%	265,32%	315,70%	514,08%	-2,21%	3,81%

Table 7. Simulation results

	Thr. Time 1		Thr. Time 2		Thr. Time 3		Average Thr. Time	
	average	St. dev	average	St. dev	average	St. dev	average	St. dev
CMS(in line)	3,27%	1,24%	2,21%	2,68%	0,42%	0,79%	2,17%	1,58%
FCMS	551,65%	775,44%	547,81%	770,79%	352,77%	508,21%	496,34%	699,14%
RCMS 7,5	76,62%	63,94%	75,71%	63,20%	-12,66%	13,99%	53,20%	43,86%
RCMS 5,3	88,43%	87,56%	87,40%	86,09%	-7,56%	5,58%	62,97%	63,25%
RCMS 3,2	113,21%	101,38%	112,44%	100,65%	17,07%	44,23%	87,72%	78,76%
RCMS 4,1	101,34%	117,21%	100,78%	116,89%	-0,88%	16,07%	74,25%	89,56%
RCMS 3,1	139,38%	162,54%	136,99%	159,71%	12,55%	31,48%	104,75%	125,62%
RCMS 2,1	208,51%	275,98%	205,19%	272,66%	38,91%	75,03%	161,62%	218,90%

Table 8. Simulation results: arrival comparison

	Inter-arrival time	WIP	Thr. time 1	Thr. time 2	Thr. time 3	Average Thr. Time	Tardiness	Throughput
CMS in line	5	3,20%	4,41%	3,21%	0,83%	3,08%	7,56%	0,17%
	6	3,08%	3,72%	4,00%	0,02%	2,94%	5,88%	0,22%
	7	2,56%	3,56%	3,51%	0,15%	2,77%	4,80%	-0,23%
FCMS	5	65,16%	77,98%	76,58%	29,02%	64,87%	247,06%	0,08%
	6	13,64%	19,60%	19,42%	-4,64%	13,72%	22,65%	-0,10%
	7	13,23%	17,97%	17,96%	0,26%	13,92%	22,87%	-0,58%
RCMS 2,1	5	18,39%	31,34%	30,25%	-18,17%	18,25%	63,03%	0,08%
	6	7,95%	14,47%	14,27%	-11,52%	8,18%	6,19%	-0,20%
	7	9,17%	13,43%	13,43%	-5,43%	9,14%	12,61%	0,12%

equal to 5) the other performance converge to values close to CMS configuration with differences about 10%. The better performance of RCMS is obtained with inter-arrival time equal to 5 therefore with a medium –high average utilization of the manufacturing system (see Table 5). With high and low congestion levels the other configurations compared to CMS have very low performance level. This is confirmed in Figure 13 that shows the profile of performance at different congestion levels.

Figures 14 and 15 show the comparison of the performance measures. It is clear that FCMS configuration in all cases performs worse especially for average inter-arrival time equal to 5. The design of this configuration needs to be rethought. For higher inter-arrival times the differences tend to decline. The behaviour of RCMS is more interesting and there is more possibility for improvement. In Figure 13 observing the curve of RCMS (2,1), it is interesting to note that the throughput time of product 3 performs better than other configurations. This is probably due to the fact that the cell 3 has lower loads (since part mix 3 is 20%) and it obtains more synergy from the remainder cell. In that configuration queues larger than 2 units (parts) are not tolerated. In this case, the remainder cell is used frequently and this is the key to a better behaviour of system configuration. The results showed indicate that a better balance of utilizations between dedicated cells and remainder cell leads to an improvement in performance.

Figure 13. Performance comparison: RCMS

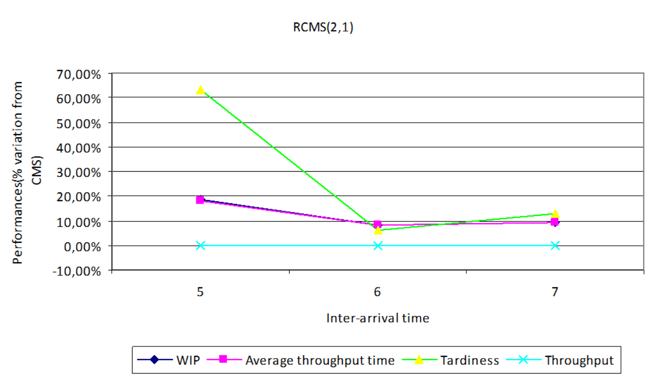


Figure 14. Performance comparison: interarrival time

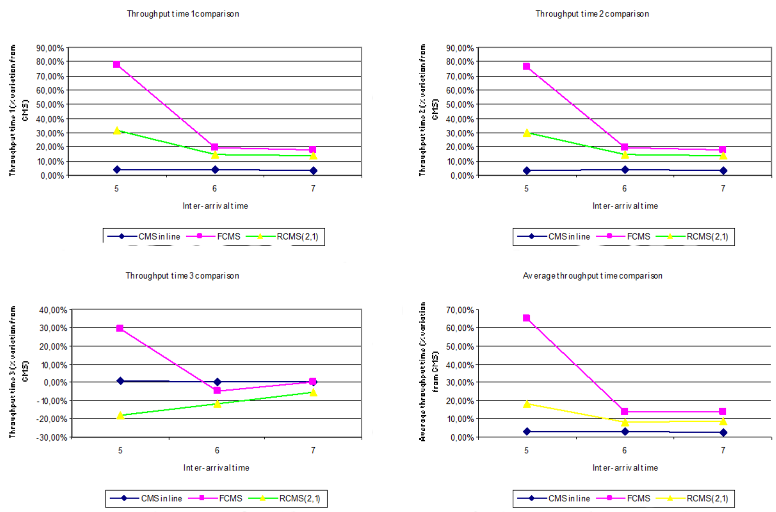
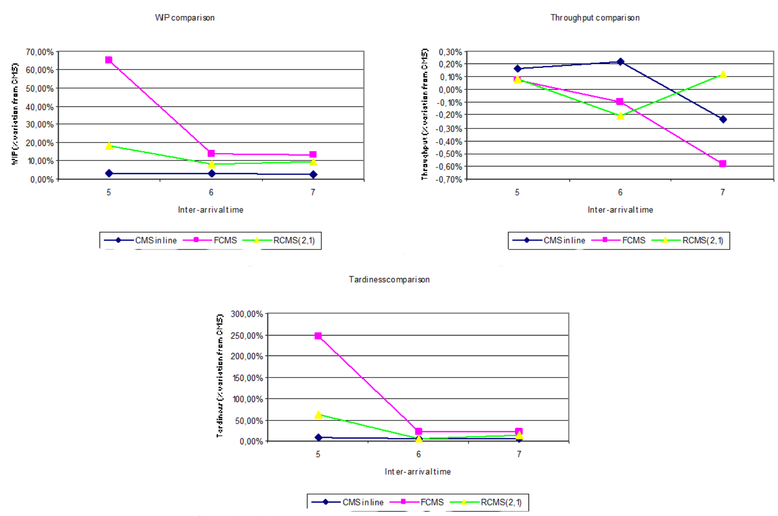


Figure 15. Performance comparison: interarrival time



CONCLUSION AND FUTURE DEVELOPMENT

This chapter investigates several configurations of the cellular manufacturing systems. A simulation environment is used to create equal operating conditions for different cellular systems. Each simulation includes the same number of machines. Thus the comparison between systems is normalized. Volume changes are analysed changing of inter-arrival times. It has been considered interesting to compare the performance because the economic environment is extremely turbulent. In particular our attention has focused on alternative approaches to traditional cells. A solution that looks interesting results is the remainder cellular manufacturing system (RCMS). The results of this research can be summarized as it follows:

- the classical cellular configuration with machines placed in line (CMS in line) is the best solution with static market conditions; the results are very close to the case of machines that are not in line (CMS);
- the fractal cellular configuration(FCMS) gives bad results as it is conceived in a static environment and should think a more complex logic with different loading policies;
- the cellular manufacturing system with remainder cell (RCMS) is already competitive in some cases with larger inter-arrival times; the best configuration is one that requires more stringent threshold values which imply a greater use of the remainder cell.

From this it follows that RCMS could become very competitive when the presence of a turbulent market would involve a greater use of remainder cell, and similarly the presence of noise on the manufacturing system(such as machine breakdowns or maintenance).

In literature, it is known that the FCMS and the RCMS are not very competitive against classical CMS.

But in previous studies remainder cells were often used as support cells with exclusive use in special circumstances. Our proposal is to adopt loading policies designed to achieve a strategic use of the remainder cells. Simple loading policies included in simulation models show how the remainder cell can be used to keep different performance under certain conditions. This work aims to demonstrate under certain dynamic conditions the proposed configurations can be competitive with classical CMS. Furthermore this chapter demonstrates the strong dependence of the results from the design of loading approaches, which deserve special attention.

Future research could focus on defining complex loading policies able to maintain high performance of the manufacturing system in different operating conditions and also taking into account the need for maintenance and possible failures of the machines(also for those belonging to remainder cell). These policies will certainly improve both the RCMS as for FCMS.

Moreover in the RCMS the logic of loading machines have a strong influence on the performance. Under dynamic conditions with market fluctuations these strategies using remainder cells can avoid the reconfigurations of manufacturing systems, avoiding downtimes and reducing costs.

Future works could investigate a variety of systems that integrate the configurations showed in this chapter with decision-making systems with intelligence to interpret the variability of real production scenario, moreover would also be interesting to analyze the economic aspect of different manufacturing solutions and how it may influence the choices.

REFERENCES

Ahkioon, S., Bulgak, A. A., & Bektas, T. (2009). Cellular manufacturing systems design with routing flexibility, machine procurement, production planning and dynamic system reconfiguration. *International Journal of Production Research*, *47*(6), 1573–1600. doi:10.1080/00207540701581809

Aramoon Bajestani, M., Rabbani, M., Rahimi Vahed, A. R., & Baharian Khoshkhou, G. (2009). A multiobjective scatter search for a dynamic cell formation problem. *Computers & Operations Research*, *36*, 777–794. doi:10.1016/j.cor.2007.10.026

Aryanezhad, M. B., & Deljoo, V., Mirzapour Al-e-hashem, S. M. J. (2009). Dynamic cell formation and the worker assignment problem: A new model. *International Journal of Advanced Manufacturing Technology*, *41*, 329–342. doi:10.1007/s00170-008-1479-4

Chen, C. H., & Balakrishnan, J. (2005). Dynamic cellular manufacturing under multiperiod planning horizons. *Journal of Manufacturing Technology Management*, *16*(5), 516–530. doi:10.1108/17410380510600491

Chen, C. H., & Balakrishnan, J. (2007). Multi-period planning and uncertainty issues in cellular manufacturing: A review and future directions. *European Journal of Operational Research*, *177*, 281–309. doi:10.1016/j.ejor.2005.08.027

Hachicha, W., Masmoudi, F., & Haddar, M. (2007). An improvement of a cellular manufacturing system design using simulation analysis. *International Journal of Simulation Modelling*, *4*(6), 193–205. doi:10.2507/IJSIMM06(4)1.089

Irani, S. A., Subramanian, S., & Allam, Y. S. (1999). Introduction to cellular manufacturing system. In Irani, S. A. (Ed.), *Handbook of cellular manufacturing systems* (pp. 29–30). John Wiley & Sons. doi:10.1002/9780470172476.ch

Kelton, W. D., & Sadowski, R. P. (2009). *Simulation with Arena*. McGraw-Hill.

Kesen, S. E., Sanchoy, K., & Gungor, Z. (2010). A genetic algorithm based heuristic for scheduling of virtual manufacturing cells (VMCs). *Computers & Operations Research*, *37*, 1148–1156. doi:10.1016/j.cor.2009.10.006

Kesen, S. E., Toksari, M. D., Gungor, Z., & Guner, E. (2009). Analyzing the behaviors of virtual cells (VCs) and traditional manufacturing systems: Ant colony optimization (ACO)-based metamodels. *Computers & Operations Research*, *36*(7), 2275–2285. doi:10.1016/j.cor.2008.09.002

Maddisetty, S. (2005). *Design of shared cells in a probabilistic demand environment.* PhD Thesis. College of Engineering and Technology of Ohio University, Ohio(USA).

Montreuil, B. (1999). Fractal layout organization for job shop environments. *International Journal of Production Research*, *37*(3), 501–521. doi:10.1080/002075499191643

Rheault, M., Drolet, J., & Abdulnour, G. (1995). Physically reconfigurable virtual cells: A dynamic model for a highly dynamic environment. *Computers & Industrial Engineering*, *29*(1–4), 221–225. doi:10.1016/0360-8352(95)00075-C

Safei, N., Saidi-Mehrabad, M., & Babakhani, M. (2007). Designing cellular manufacturing systems under dynamic and uncertain conditions. *Journal of Intelligent Manufacturing*, *18*, 383–399. doi:10.1007/s10845-007-0029-5

Safei, N., & Tavakkoli-Moghaddam, R. (2009). Integrated multi-period cell formation and subcontracting production planning in dynamic cellular manufacturing systems. *International Journal of Production Economics*, *120*, 301–314. doi:10.1016/j.ijpe.2008.12.013

Selim, M. S., Askin, R. G., & Vakharia, A. J. (1998). Cell formation in group technology: Review evaluation and directions for future research. *Computers & Industrial Engineering*, *34*(1), 3–20. doi:10.1016/S0360-8352(97)00147-2

Süer, G. A., Huang, J., & Maddisetty, S. (2010). Design of dedicated, shared and remainder cells in a probabilistic demand environment. *International Journal of Production Research*, *48*(19), 5613–5646. doi:10.1080/00207540903117865

Vakharia, A. J., Moily, J., & Huang, Y. (1999). Evaluating virtual cells and multistage flow shops: An analytical approach. *International Journal of Flexible Manufacturing Systems*, *11*, 291–314. doi:10.1023/A:1008117329327

Venkatadri, U., Rardin, R. L., & Montreuil, B. (1997). A design methodology for fractal layout organization. *IIE Transactions*, *29*, 911–924. doi:10.1080/07408179708966411

Wang, X., Tang, J., & Yung, K. (2009). Optimization of the multi-objective dynamic cell formation problem using a scatter search approach. *International Journal of Advanced Manufacturing Technology*, *44*, 318–329. doi:10.1007/s00170-008-1835-4

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 366-384, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 31
Optimization and Mathematical Programming to Design and Planning Issues in Cellular Manufacturing Systems under Uncertain Situations

Vahidreza Ghezavati
Islamic Azad University, Iran

Mohammad Saeed Jabal-Ameli
University of Science and Technology, Iran

Mohammad Saidi-Mehrabad
University of Science and Technology, Iran

Ahmad Makui
University of Science and Technology, Iran

Seyed Jafar Sadjadi
University of Science and Technology, Iran

ABSTRACT

In practice, demands, costs, processing times, set-up times, routings, and other inputs to classical cellular manufacturing systems (CMS) problems may be highly uncertain, which can have a major impact on characteristics of manufacturing system. So, development models for cell formation (CF) problem under uncertainty can be a suitable area for researchers and belongs to a relatively new class of CMS problems that not researched well in the literature. In this way, random parameters can be either continuous or described by discrete scenarios. If probability information is known, uncertainty is described using a (discrete or continuous) probability distribution on the parameters, otherwise, continuous parameters are normally limited to lie in some pre-determined intervals. This chapter introduces basic concepts about uncertainty themes associated with cellular manufacturing systems and briefly reviews literature survey for this type of problem. The chapter also discusses the characteristics of different mathematical models in the context of cellular manufacturing.

DOI: 10.4018/978-1-4666-1945-6.ch031

INTRODUCTION

During the past few decades, there have been various types of optimization techniques and mathematical programming approaches for cellular manufacturing systems under different random situations. In a cell manufacturing, once work cells and scheduling of parts in each cell are determined, it may be possible that cycle time in a specific cell be more than the other cells which creates a bottleneck in a manufacturing system. In this way, there are two different approaches in order to decrease cycle time in bottleneck cell: duplicating bottleneck machines or outsourcing exceptional parts which are known as group scheduling (GS) in the literature. Selecting each approach to balance cycle times among all cells can lead to variations in machines layout characteristics by changes in type and number of machines. Finally, formations of cells are also changed according to the changes in scheduling decisions. Thus, scheduling problem is one of the operational issues which must be addressed in design stage concurrently in an integrated problem so that the best performance of cells would be achieved. It is noted that scheduling problem includes many tactical parameters with random and uncertain characteristics. In addition, uncertainty or fluctuations in input parameters leads to fluctuations in scheduling decisions which could reduce the effects of cell formation decisions. Figure 1 indicates transmission of uncertainty from tactical parameters to the CMS problem.

Thus, in order to intensify effectiveness of the solution, integrated problem in uncertain conditions must be studied so that final solution will be robust and immune against the fluctuations in input parameters.

In the concerned problem, uncertain parameters can be listed as follows:

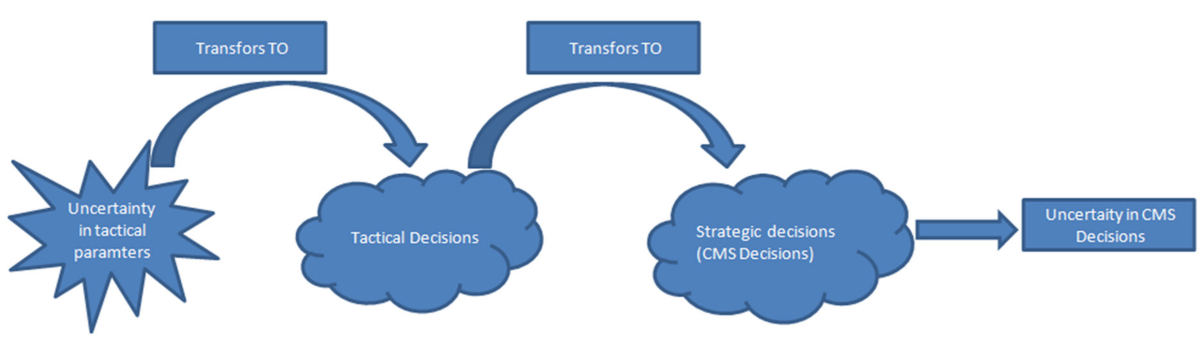
- Demand,
- Processing time,
- Routings or machine-part matrix,
- Machines' availability,
- Failure rate of machines,
- Capacities,
- Lead times,
- Set-up considerations,
- Market aspects,
- …,

where the impact of each factor is discussed in the following sections.

PROBLEM BACKGROUND

Group technology (GT) is a management theory that aims to group products with similar processes or manufacturing characteristics, or both. Cellular manufacturing system (CMS) is a manufacturing

Figure 1. Illustration of uncertainty transmission to the CMS decision



concept to group products into part families based on their similarities is manufacturing processing. Machines are also grouped into machine cells based on the parts which are supposed to be manufactured by these machines. CMS framework is a important application of group technology (GT) philosophy. The basic purpose of CM is to identify machine cells and part families concurrently, and to assign part families to machine cells in order to minimize the intercellular and intracellular costs of parts. Some real-world limitations in CF are:

- Available capacity of machines must not be exceeded,
- Safety and technological necessities must be met,
- The number of machines in a cell and the number of cells have not be exceeded an upper bound,
- Intercellular and intracellular costs of handling material between machines must be minimized,
- Machines must be utilized in effect (Heragu, 1997).

Aggregating traditional considerations with newly ones such as scheduling, stochastic approaches, processing time, variable demand, sequencing, and layout consideration can be more practical. This survey highlights studies that are relevant to the uncertainty planning of CMS problems; however, a survey of certain conditions will also be presented.

Cellular manufacturing decisions are strategic decisions which can be affected by operational decisions such as scheduling, production planning, layout consideration, utilities, productivity and etc. Thus, in order to effecting decision making related to cell formation design, it is necessary to integrate strategic decisions and operational decisions in a single problem. Recently, researchers have had some efforts in order to integrate two types of decisions. But the lack of literature is that most of them are studied in certain situations while in real-world most of the operational parameters are uncertain; and thus, integrated problems must be more studied in uncertain situations.

In the literature correspondence to CMS problems, uncertainty has been considered under different circumstances. We have classified previous researches into different groups which are discussed next.

Group 1: Uncertainty could appear either in demand or in products' mix. In this group, there are two approaches of fuzzy theory and stochastic optimization to handle uncertainty. In some of them stochastic demand is aggregated with tactical aspects such as production planning (Hurley and Whybark 1999), layout problem (Song and Hitomi1996) or dynamic and multi period conditions (Balakrishnan and Cheng 2007). Also, in other studies, uncertainty in products' demand has been resolved by fuzzy approach (Safaei et. al. 2008).

Group 2: Researchers formulated and analyzed CMS problem considering fuzzy coefficients in the objective function and constraints (Papaioannou and Wilson 2009).

Group 3: Processing times of products are assumed to be uncertain where mathematical programming and fuzzy approaches are implemented to obtain the results which are immune against the perturbation on the uncertainty. Also, some studies such as Sun and Yih (1996) and Andres et. al. (2007) attempted to achieve solutions by heuristic procedures. Some studies have formulated the problem as a queue network and then analyzed it by queuing theory (Yang and Deane 1993).

Group 4: Uncertainty normally appears due to fluctuations in design aspects during production process. Since, fluctuations in design aspects are not certain events, so uncertainty can be formulated by a set of future scenarios. In this way, some studies applied interval coefficient to resolve uncertainty (Shanker and Vrat 1998).

Group 5: In some explorations, uncertainty has been considered in the availability of resources for production equipments. In this way, some works have formulated CMS problem applying probability theory (Kuroda and Tomita (2005) and Hosseini 2000). In addition, some of them considered multi processing routes to be substituted once a machine encounters with failure (Siemiatkowski and Przybylski (2007) and Asgharpour and Javadian 2004).

Group 6: Uncertainty has been recognized in similarity coefficients. For example, a new similarity coefficient has been introduced where applied fuzzy theory and then transformed it to a binary matrix (Ravichandran and Chandra Sekhara Rao 2001).

Group 7: Capacity level of machines is considered to be uncertain. Since this critical parameter has an important role to determine bottleneck machine, thus it is vital to make flexible decisions under any realization of this parameter (Szwarc et al 1997).

Group 8: Finally, uncertainty in CMS problem has been detected in products arrival time to cells. Classical models assume that all products are available at the beginning of the production planning while in real application it may be occurred that products arrive to cell with unknown time. In this way, researchers modeled CMS problem as queue network to resolve uncertainty (Yang and Deane 1993).

Literature survey classifications can be described as follows. There exist many researches in certain situations for designing CMS in different areas such as cell formation integrated with scheduling (Solimanpur et al. (2004), Aryanezhad and Aliabadi et al 2011), considering exceptional elements in CF (Tsai et al. (1997), Mahdavi et al. 2007), some works apply meta-heuristics and heuristics methods to solve large scale problems are more practical and appealing real-case problems (Xiaodan Wu et al (2006), Venkataramanaiah 2007).

OPTIMIZATION APPROACHES IN UNCERTAIN SITUATIONS

Rosenhead et al (1972) divided decision environments into three groups of deterministic, risk and uncertain. In deterministic situations, all problem parameters are considered to be given. In risk problems parameters have probability distribution function where it is known for decision maker while in uncertain situations there is no information about probabilities.

The problems which are classified into the risk are named stochastic and the primary objective is to optimize expected value of system outcome. Also, the uncertain problems are known as robust and the primary objective is mainly to optimize performance of the system in the worst case conditions.

The aim of both stochastic and robust optimization methods is to find solution with a suitable performance in realization of any value for uncertain parameter.

Random parameters can be either continues or explained by discrete scenarios. If probability information are known, uncertainty will be explained by continues or discrete distribution functions. But if no information is available, parameters are assumed to be in predefined intervals. Scenario planning is a method in which decision makers achieve uncertainty by indicating a number of possible future states. In such conditions, the goal is to find solutions which perform well under all scenarios. In some cases, scenario planning replaces predicting as a way to assess trends and potential modifications in the industry environment (Mobasheri et al 1989). Decisions makers can thus develop strategic responses to a range of environmental adjustments, more adequately preparing themselves for the uncertain future. Under such conditions, scenarios are qualitative descriptions of possible future states, consequences from the present state with consideration of potential key industry events. In other cases, scenario planning is used as a tool for modeling and solving specific operational problems (Mulvey 1996). While scenarios here also depict a range of future states, they do so through quantitative descriptions of the various values that problem input parameters may resolve. Scenario based planning has two main negative aspects. The first is that identifying scenarios and assigning probabilities to them are a difficult task. The second is that we are unable to increase the number of scenarios since due to limitation on computation time which consequently limits the future correspondence situations for decision making. This approach has the advantageous that provides statistical correlation between parameters (Snyder 2006).

DECISION MAKING APPROACHES IN UNCERTAIN SITUATIONS

There are different approaches which can be applied in modeling process based on the problem characteristics: Stochastic Optimization (SO), Robust Optimization (RO) and Queuing Theory (QT) with defined decision tree as follows.

- Stochastic Optimization
- Discrete Planning - Set of Scenario
- Continues Optimization
- Mean Value model: the most popular objective in any SO problem is to optimize expected value of the system outcome. For example, minimizing expected cost or maximizing expected income.
- Mean – Variance Model: in some studies variance and expected of system performance are considered simultaneously in optimization problem.
- Probability Approaches
- Max Probability Optimization: Maximizing the probability of a random event that solution performs good under each realization of random parameter.
- Chance Constrained Programming: a probability event located in problem constraint sets such as service level constraint.
- Queuing Theory & Markov Chain: It is a well-known approach.
- Robust Optimization

The objective in any stochastic optimization problem mainly focuses on optimizing the expected value of system outcome such as maximizing expected profit or minimizing total expected cost.

In any stochastic programming we must determine which variables are considered in the first stage (design variable) and which are considered in the second stage (control variable). In other words, which variables must be determined first and which of them must be determined after uncertainty is resolved. In modeling process for cellular manufacturing problem, cell formation decisions are the first and operational and tactical decisions are the second variables. If both decisions are made in a single stage, the model is reduced to a certain problem in which uncertainty of parameters are replaced by mean of variables.

Mean-Variance Models

The mean value models discuss only the expected performance of the system without reflecting on the fluctuations in performance and the decision maker's risk aversion limitations. However, a portion of literature incorporates the company's level of risk aversion into the decision-making process, classically by applying a mean–variance objective function.

$$\text{Min} = E(\text{Cost}) + \lambda \text{Var}(\text{Cost})$$

Probabilistic Approaches

The mean-variance models consider only the expected value or variance of the stochastic objective function, there is an extensive portion of literature which considers probabilistic information about the performance of the system; for example, maximizing the probability that the performance is good or minimizing the probability that it is bad, under suitable and predefined explanations of "good" and "bad". We introduce two such approaches: (1) max-probability problems; (2) chance-constrained programming;

Queuing Theory for CMS Problem

Queuing theory can be applied to any manufacturing or service systems (also, in cellular manufacturing systems). For example, in a machine shop, jobs wait to be machined; (Heragu 1997b). In a queuing system, customers arrive by some arrival process and wait in a queue for the next available server. In the manufacturing framework, customers can be assumed as parts and servers may be machines or working cells. The input process shows how parts arrive at a queue in a cell. An arrival process is commonly identified by the probability distribution of the number of arrivals in any time interval. The service process is usually described by a probability distribution. The service rate is the number of parts served per unit time. The arrival rate of a queuing system is usually given as the number of parts arriving per unit time. Thus, measurements of a queue system such as maximization the probability that each server is busy (utilization factor), minimization waiting time in queues (that leads to minimization work in process in cells) and etc can be optimized and cells will be formed optimality.

Robust Optimization

Once there is no probability information about the uncertain parameters, the expected cost and other objectives discussed in previous section are inappropriate. Many measurements of robustness have been introduced for this condition. The two most common are mini-max cost and mini-max regret, which are directly related to one another. Just like the stochastic optimization case, uncertain parameters in robust optimization problems may be considered as being either discrete or continuous. Discrete parameters are formulated applying the scenario based planning. Continuous parameters are normally assumed to lie in some predefined interval, because it is often impossible to consider a "worst case scenario" when parameter values are unbounded. This type of uncertainty is described as "interval uncertainty".

The two most common robustness measurements consider the regret of a solution, which is the difference (absolute or percentage) between the cost of a solution in a given scenario and the cost of the optimal solution for that scenario. Regret is sometimes described as opportunity loss: the difference between the quality of a given strategy and the quality of the strategy that would have been chosen had one known what the future held (Snyder 2006).

As it was already described, the performance of a cellular manufacturing system heavily influenced by tactical and operational decisions such as scheduling, production planning, layout and etc. Notable point is that the tactical decisions and operational parameters are dependent on many uncertainties that affect the system. As a result, the tactical and operational decisions are suffering from uncertainty. This causes to transfer uncertainty into the decisions related cell formation. Therefore, it is essential for researchers to recognize different types of uncertainty in the problem and make decisions regarded to their impact into the problem.

The most important parameters with uncertainty in manufacturing cell formation problem considered as below:

- Demand

- Processing time
- Routings or machine-part matrix
- Machine's failure rate
- Capacities

One of the factors causing uncertainty in the problem associated with product design changes during the course of production. Moreover, changes in product design with many features of the product are altered. Design changes can occur based on a variety of reasons such as changes in customer expectations, short-term life products, and competing products to market entry.

Under such circumstances, many characteristics of products such as demand and time will find a process of change.

Note that the reasons of changes expressed are not certain events in the future and thus they have to be predicated as some discrete scenarios. In such case analytical space problem is discrete and can be optimized by discrete optimization.

As it was discussed earlier, one of the product features which can be changed due to changes in product design is product routings. In this way, sequence of machines in which product has to visit them may be changed and therefore part – machine index may be changed. In such cases, the values within the part – machine matrix unlike classical models that were only zero or one can be a probabilistic value between zero and one. In such problems, discrete optimization can be applied to formulation.

Another factor with uncertainty is the rate of access to machines based on their failure. Since failure and machine downtime are not certain events, the machine accessibility for the decision maker at the time of manufacturing cells with defined uncertainty is also under uncertainty.

Another parameter that is uncertain and can affect formation of work cells is features of capacity. These factors include different items: the capacity of processing machinery on parts as well as physical capacities for manufacturing framework. Such variations must be predicted at the beginning planning horizon.

The summary of above discussions can be found Table 1.

MATHEMATICAL MODELLING

In this section, different mathematical models with different optimization approaches which include two new models and one published model are discussed. The selected approaches are stochastic optimization and queuing theory.

Table 1. Summary uncertainty developments in CMS problem

No.	Uncertain parameter	Optimization Approach	Decision space
1	demand	Stochastic	Continuous & Discrete
2	Processing time	Stochastic	Continuous & Discrete
3	Processing time	Robust	Continuous & Discrete
4	Processing time	Queuing Theory	Continuous
5	Routing	Stochastic	Discrete
6	Routing	Queuing Theory	Discrete
7	Capacity	Stochastic	Discrete
8	Machines' Availability	Queuing Theory	Continuous & Discrete
9	Machines' Availability	Stochastic	Continuous & Discrete
10	Lead times	Stochastic & Robust	Continuous & Discrete

Model 1

In this section, a bi-objective mathematical model to form manufacturing cells is presented where uncertainty is accessed in part – machine matrix. As discussed earlier, due to changes in design characteristics of products, several factors are subject to changes such as the processing routings of parts. Thus, according to the forecasting based on the scenario planning, forecasting different routing processes for a part in uncertain situation is possible. In this condition, each part can have different routing process for each scenario. Therefore, in order to design cellular configuration efficiently, all planning conditions must be considered. In current problem the factor with uncertainty is part – machine matrix. In classical models, only zero-one elements are used in part – machine matrix while in the presented problem each element can be a continuous value between zero and one. Each array denotes the probability that part i visits machine j with regard to all scenarios. For example, if there are two scenarios in which the probability of the first scenario is 0.4 and for the second one is 0.6, we have:

$p_1 = 0.4 \Rightarrow$ Routing in scenario 1 for part 1: Machine 1 → Machine 2 → Machine 3 → Finish

$p_2 = 0.6 \Rightarrow$ Routine in scenario 2 for part 1: Machine 1 → Machine 2 → Machine 3 → Finish

$$a_{[ij]} = \begin{bmatrix} M.1 & M.2 & M.3 & M.4 \\ 1 & 0.4 & 1 & 0.6 \end{bmatrix}$$

Where element [ij] indicates the probability that part i processed on machine j.

Since, in both scenarios, machines 1 and 3 are the same in processing routing, so part 1 has to visit them surely (or with probability 1) to do operation process. But, based on the first scenario this part has to visit machine 2 with probability 0.4 and also machine 4 with probability 0.6. As it can be seen, in introduced part – machine matrix each array can have a value between zero and one based on the probability occurrence for scenarios.

In a mathematical model which is presented in this section, the first objective function minimizes the costs associated with the under utilization in a manufacturing system. Also, the second objective function is optimizing a random event in manufacturing system unlike the classical models which optimized only certain events. As it was discussed in definitions of a cellular manufacturing system, one of the most important objectives is to minimize the number of inter cellular transportation. In this problem, since processing rout for parts is uncertain, therefore the number of inter cellular transportation is uncertain too. A random event which is considered for optimization is to "minimizing the probability that the number of inter cellular transportation exceeds the upper bound limitations". For computing above objective the following notations are defined:

Parameters:

$$a_{ijs} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if part *i* needs to be processed on machine *j* in scenario *s*
0 otherwise

p_s: Probability of occurring scenario s
N: Maximum number of intercellular transportation allowed in each scenario

Decision Variables:

n_s: Number of intercellular transportation in scenario *s*.

$$e_s = \begin{cases} 1 \\ 0 \end{cases}$$

1 if no. of intercellular transportation in scenario *s* configuration exceeded up bound *N*
0 otherwise

or

$$e_s = \begin{cases} 1 \\ 0 \end{cases}$$

1 if $n_s \geq N$
0 if $n_s < N$
z_s: Integer additional variable for each scenario.

$$x_{ik} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if part *i* is assigned to cell *k*
0 otherwise

$$y_{jk} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if machine *j* is assigned to cell *k*
0 otherwise

In order to minimizing under utilization costs in the first objective function, the following function is defined:

$$MinZ_1 = \sum_s \sum_i \sum_j p_s \times (1 - a_{ijs}) \times x_{ik} \times y_{jk} \tag{1}$$

Also, based on the above definitions, an attractive random event for minimizing the second function can be defined as follows:

p (no. of intercellular transportation in each condition $\geq N$)

The above random event must be optimized by minimizing the probability of occurrence that leads to maximum utility for decision maker in final solution. In other words, above probability transforms to the following function:

$$MinZ_2 = \sum_s e_s \times p_s \tag{2}$$

Since there is s scenarios in the proposed problem which are similar to s independent random events, thus probability of total events will be equal to summation of probability of each event. In other words, assuming $s_1, s_2, \ldots, s_n$ as *n* independent random events, we have:

$$P\left(s_1 \cup s_2 \cup \ldots \cup s_n\right) = P\left(s_1\right) + P\left(s_2\right) + \ldots + P\left(s_n\right)$$

As a result, in above function if in scenario s the number of inter cellular transportation exceeds the upper bound limitation then we can assume that inter cellular transportation may be occurred with the probability of p_s. Finally, the summation of the probability of scenarios with unsatisfied inter cellular transportation restriction denotes the final probability of the problem.

In this model, the objective functions and also, the following constraints are effective:

$$MinZ_1 = \sum_s \sum_i \sum_j p_s \times (1 - a_{ijs}) \times x_{ik} \times y_{jk}$$
$$MinZ_2 = \sum_s e_s \times p_s$$

Constraints:

$$\sum_k x_{ik} = 1 \quad \forall i \tag{3}$$

$$\sum_k y_{jk} = 1 \quad \forall j \tag{4}$$

$$n_s - \sum_i \sum_j a_{ijs} \times x_{ik} \times (1 - y_{jk}) = 0 \tag{5}$$

$$z_s - \left[\frac{n_s}{N}\right] = 0 \qquad (6)$$

$$z_s \leq M \times e_s \qquad (7)$$

$$x_{ik}, y_{jk}, e_s \in \{0,1\} \quad z_s \text{ integer} \geq 0 \quad n_s \geq 0$$

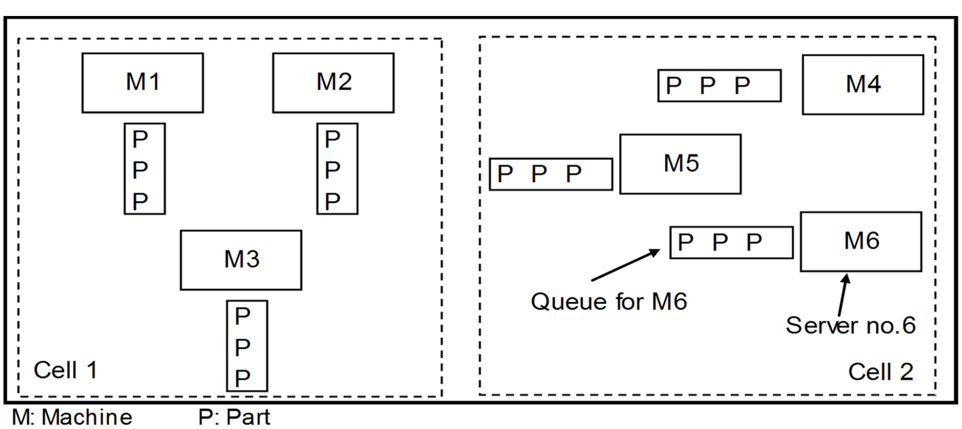
The first objective minimizes total expected cost associated with the utilization computed when a part do not need to be processed on a machine placed together in a same cell. The second objective minimizes the probability that number of inter cellular transportation exceeds the maximum transportation. Set constraint (3) says that each part must be assigned to a single cell. Set constraint (4) states that each machine can be assigned only to one cell. Set constraint (5) computes total number of inter transportation in each scenario. In set constraint (6) additional variable z_s will be zero if the number of inter transportations in scenario s is less than the maximum limit and it is an integer value greater than 1 else. Set constraint (7) guarantees that if $n_s \geq N$ then e_s will be 1. Otherwise, e_s will be 0.

Model 2

Applying Queuing Theory to CMS Problem

In this section, we formulate a CMS problem as a queue system. Also, assume a birth-death process with constant arrival (birth) and service completion (death) rates. The role of the birth-death process in automated manufacturing systems is described in detail in Viswanadham and Narahari (Viswanadham and Narahari 1992). Specifically, let λ and μ be the arrival and service rate of parts, respectively, per unit time. If arrival rate is greater than the service rate, the queue will grow infinitely. The ratio of λ to μ is named utilization factor or the probability that a machine is busy and is defined as $\rho = \lambda \mid \mu$. *T*herefore, for a system in steady state, this ratio must be less than one. In this research, we assume M/M/1 queue system for each machine in CMS where each part arrives to cells with rate *λ*i *an*$_d$ parts served by machines. In this condition, due to operate different parts (or different customers) on each machine and each part has different arrival rate, so for each machine (server) ρ is computed using the following property. Figure 2 illustrates modeling of cellular manufacturing system by queuing theory approach.

Figure 2. A CMS problem and queuing theory framework (Ghezavati and Saidi-Mehrabad 2011)



Pr**operty 1 t**he minimum of independent exponential random variables is also, exponential. Let F1, *F2, ..., Fn* be independent random variables with parameters $\lambda 1$, $\lambda 2, ..., \lambda n$. Let Fmin = min{F1, F2, ..., Fn}. Then for any t ≥ 0,

$$P(F_{\min} > t) = P(F_1 > t) \times P(F_2 > t) \times ... \times P(F_n > t)$$
$$= e^{-\lambda_1 t} e^{-\lambda_2 t} ... e^{-\lambda_n t} = e^{-[\lambda_1 + \lambda_2 + ... + \lambda_n]t}$$

An interesting implication of this property to inter-arrival times is discussed in Hillier and Lieberman (Hillier and Lieberman 1995). Suppose there are n types of customers, with the ith type of customer having an exponential inter-arrival time distribution with parameter λi, arrive at a queue system. Let us assume that an arrival has just taken place. Then from a no-memory property of exponential distribution, it follows that the time remaining until the next arrival is also exponential. Using mentioned property, we can see that the inter-arrival time for entire queue system or efficient arrival rate (which is the minimum among all inter-arrival times) has an exponential distribution with parameter:

$$\lambda_{eff} = \sum_{i=1}^{N} \lambda_i$$

Hence, utilization factor or the probability that each machine (j) is busy is as follow (efficient arrival rate divided by service rate):

$$\rho_j = \frac{\lambda_{eff}}{\mu_j} = \frac{\sum_{i=1}^{N} \lambda_i}{\mu_j} \tag{8}$$

Chance Constrained Programming

Since, both arrival time and service time are uncertain so the amount of time in which each customer spends in server will be uncertain, too. In order to prevent long waiting time for each customer, a chance constraint must be considered in the formulation. Note that distribution function denoting total time for each customer in a M/M/1 system is as follows:

$$P(W_s \geq t) = e^{-\mu(1-\rho)t} \tag{9}$$

Proof: Assume that there are N customers in a system once a new customer is arrived. Thus, based on the conditional probability theory:

$$P(W_s \geq t) = \sum_{n=0}^{\infty} P(W_s \geq t \mid N = n) \times P(N = n) \tag{10}$$

On the other side, total time in which a new customer has to wait is equal to:

$$W_q = F_1 + F_2 + ... + F_n \tag{11}$$

where F_i denotes service time for customer i. So:

$$W_s = W_q + F_{n+1} \tag{12}$$

where F_{n+1} denotes service time for new arrived customer. It is obvious that sum of the n+1 random variables with exponential distribution with rate μ will be an Erlang random variable with parameters n+1 and μ. So:

$$P(W_s \geq t \mid N = n) = P(\sum_{i=1}^{n+1} F_i > t)$$
$$= \int_t^{\infty} \mu \times e^{-\mu \cdot y} \frac{(\mu y)^n}{n!} d_y \tag{13}$$

Note that the probability of being n customers in a M/M/1 model system is:

$$p_n = \rho^n (1-\rho) \quad \text{where} \quad \rho = \frac{\lambda}{\mu} \tag{14}$$

Based on the Equations 13 and 14, Equation 10 will be computed as:

$$P(W_s \geq t) = \sum_{n=0}^{\infty} \rho^n (1-\rho) \int_t^{\infty} \mu \times e^{-\mu \cdot y} \frac{(\mu y)^n}{n!} d_y \tag{15}$$

$$= \mu(1-\rho) \int_t^{\infty} e^{-\mu \cdot y} \sum_{n=0}^{\infty} \frac{\rho^n (\mu y)^n}{n!} d_y \tag{16}$$

Also, based on the exponential series, we have:

$$\sum_{n=0}^{\infty} \frac{\rho^n (\mu y)^n}{n!} = e^{\rho \cdot \mu \cdot y} = e^{\lambda \cdot y} \tag{17}$$

If we replace Equation 17 to the Equation 16, the Equation 9 will be proven. It can be found that W_s has an exponential distribution function with parameter $\mu - \lambda$.

In order to satisfy service level this probability must be at most α. So, the chance constraint will be determined as $P(Ws \geq t) \leq \alpha$. In order to linearize this nonlinear constraint the following procedure is performed:

$$P(W_s \geq t) \leq \alpha \tag{18}$$

$$\Rightarrow e^{-\mu(1-\rho)t} \leq \alpha \tag{19}$$

$$\Rightarrow -\mu(1-\rho)t \leq Ln(\alpha) \tag{20}$$

The achieved constraint indicates that a customer will be in system more than critical time t with probability at most α.

Property 2. If n types of customers have to visit a server to receive service with different arrival rate λi then the probability of a random customer in which visits the server be ith type will be as follows:

$$p_i = \frac{\lambda_i}{\sum_j \lambda_j} \tag{21}$$

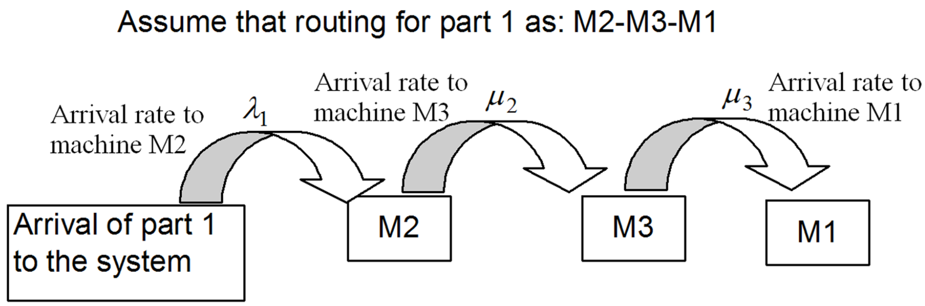
In concerned model, the characteristics of a Jacson service network will be applied. In a Jacson network, it is assumed that each customer has to visit multiple servers in order to complete service stages. For example, each part refers to several machines to complete operation processes. In such network, input rate for machines needed for the first operation will be equal to the arrival rate of the part to the system. But, the ratio for any machines need for the second operation input rate will be equal to the output rate from the previous server (or machine). Similarly, any machine needs for the third operation input rate will be equal to the exit rate of the second machine and this process goes on for the other machines.

In a cellular manufacturing problem formulated as a queue system, each part based on its routing process visits machines or multi cells in order to receive service. Figure 3 illustrates such process.

For each machine, effective input rate is made of two elements. The first fraction is the summation of arrival rate of parts which visit the machine in the first operation. The second fraction is the summation of input rate of parts which visit the machine after the second operation. This rate is equal to the output rate of the previous machine. Figure 3 illustrates difference between arrival rates for machines per a specific part. In this model, such procedure will be applied to compute effective input rate for each machine.

In this section, a part—machine matrix—will be applied where sequence operations of parts are determined. This can help us formulate problem as a Jacson network. Each element of this matrix is defined as follows:

Figure 3. Arrival rate for part 1 into the different machines based on the routing



$$a_{ik} = \begin{cases} j \\ 0 \end{cases}$$

j if kth process of part i is completed by machine j
0 otherwise

$$b_{ij} = \begin{cases} k \\ 0 \end{cases}$$

k if part i refers to machine j to complete kth process
0 otherwise

Other parameters are defined as follows:

$$z_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if operation on machine j is the first operation of part i
0 otherwise

$$c_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if part i needs to be processed on machine j
0 otherwise

λI=Arrival rate of part i to manufacturing system.
μj = service rate of machine j (or [$1/\mu j$] denotes average operation time on machine j).
pi_j = The probability that a random part type i leaves machine j.
β = Penalty rate multiplied to arrival process if intercellular movement occurs.

It is assumed that if an operation of a part has to transfer to the other cell (or inter cellular movement) then arrival rate will be multiplied by β in which included transfer time and also waiting time between cells.

λj_e^{ff} = Effective arrival rate for machine j.

Based on the above definitions λj_e^{ff} will be computed by the following equation.

$$\lambda_j^{eff} = \sum_{i=1}^{m}\left(z_{ij} \times \lambda_i\right) + \sum_{i=1}^{m}(1 - z_{ij}) \times c_{ij} \times \lambda^{eff}_{a_{i,b_{ij}-1}} \times p_{i,a_{i,b_{ij}-1}}$$

In above equation, in order to compute effective input rate for each machine two fractions are considered: the first fraction is the summation of the arrival rate for parts which visit machine j in the first operation. The second term is the summation of input rate for parts which visit the machine after the second operation. Number of operation which completed by machine j is b_{ij} based on the defined parameters. Thus, number of previous operation is $b_{ij} - 1$. Finally, according to the definition of a_{ik} (the machine completes kth operation of part i), the machine which completes previous operation of part i will be $a_{i,b\mu-1.}$ Therefore, the second term of above equation effective arrival rate for parts visit machine j after the second operation are computed as follows: effective arrival rate of a machine needs before machine j multiplied by the probability of leaving for part i from previous machine.

For example, assume that customers arrive to a book store with Poisson distribution with rate 10 per hour where 60 percent is man and 40 percent is women. Hence, the number of men arrives to the store will be Poisson with rate 15×0.6 per hour and also, number of women arrives to the store will be Poisson with rate 15×0.4 per hour.

Note that if operation j of part i needs inter cellular transportation, machine j is penalized by increasing arrival rate of part i and the rate is multiplied by β. Finally, the model must determine whether each operation needs inter cellular transportation or not. It must be mentioned that operation j of part i needs inter cellular transportation when machine j and part i are not located in the same cell. Based on the above description λeff^{i}s computed as Equation 22.

Mathematical Model

In this section, a mathematical model optimizes cell formation decisions based on the queuing theory will be proposed. The objective function is to minimize total cost included under utilization cost. Also, a chance constraint will be considered in order to prevent additional waiting time of parts in a queue line in front of each machine. As it was discussed, assuming each machine as a M/M/1 model, the chance constraint (13) satisfies considered objective.

$$MinZ = \sum_{i}\sum_{j}(1 - a_{ij}) \times x_{ik} \times y_{jk} \qquad (23)$$

Constraints:

$$\sum_{k} x_{ik} = 1 \quad \forall i \qquad (24)$$

$$\sum_{k} y_{jk} = 1 \quad \forall j \qquad (25)$$

$$\rho_j - \frac{\lambda_j^{eff}}{\mu_j} = 0 \qquad (27)$$

$$-\mu_j \times (1 - \rho_j)t \leq Ln(\alpha) \quad \forall j \qquad (28)$$

$$\rho_j \leq 1 \quad \forall j \qquad (29)$$

$$p_{ij} - \frac{\lambda_i \times c_{ij}}{\sum_{r=1}\lambda_r \times c_{rj}} = 0 \quad \forall i,j \qquad (30)$$

$$x_{ik}, y_{jk} \in \{0,1\} \quad \rho_j, p_{ij} \geq 0$$

Equation 22.

$$\lambda_j^{eff} = \sum_{i=1}^{m}\left[\left(x_{ik} \times y_{jk}\right) \times \lambda_i + x_{ik} \times \left(1 - y_{jk}\right) \times \lambda_i \times \beta\right] \times z_{ij}$$
$$+ \sum_{i=1}^{m} p_{i,a_{i,b_{ij}-1}} \times \left[\left(x_{ik} \times y_{jk}\right) \times \lambda_{a_{i,b_{ij}-1}}^{eff} + x_{ik} \times \left(1 - y_{jk}\right) \times \lambda_{a_{i,b_{ij}-1}}^{eff} \times \beta\right] \times c_{ij} \times (1 - z_{ij})$$

Equation 26.

$$\lambda_j^{eff} = \sum_{i=1}^{m} \left[\left(x_{ik} \times y_{jk} \right) \times \lambda_i + x_{ik} \times \left(1 - y_{jk} \right) \times \lambda_i \times \beta \right] \times z_{ij} + \sum_{i=1}^{m} p_{i,a_{i,b_{ij}-1}} \times \left[\left(x_{ik} \times y_{jk} \right) \times \lambda^{eff}_{a_{i,b_{ij}-1}} + x_{ik} \times \left(1 - y_{jk} \right) \times \lambda^{eff}_{a_{i,b_{ij}-1}} \times \beta \right] \times c_{ij} \times (1 - z_{ij})$$

Constraints (24) and (25) compute effective arrival rate and utilization factor for each machine, respectively. Set constraint (28) guaranties satisfaction of chance constrained for each machine where the probability that each part has to wait more than critical time *t* is at most a. Set constraint (29) ensures that utilization factor for each machine will be less than one. Set constraint (30) determines the probability that a random part leaves machine *j* be type *i*.

Model 3

Recently, Ghezavati and Saidi-Mehrabad (2010) proposed a stochastic cellular manufacturing problem in where uncertainty is captured by discrete fluctuations in processing times of parts on machines. The aim of their model was to optimize scheduling cost (expected maximum tardiness cost) plus cell formation costs, concurrently. The mathematical model is represented in this part and interested readers are referred to read the paper for more details.

Parameters:

$$a_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if part *i* required to be process on machine *j*
0 otherwise

c_i: *Penalty* cost of subcontracting for part *i*

$\mu i_{j:}$ Cost part i not utilizing machine j

M_{max}: Maximum number of machines permitted in a cell

$C_{\mu:}$ Maximum number of cells permitted

p_s: Probability of scenario *s* occurs

t_{ijs}: Processing time for part *i* on machine *j* in scenario *s*

DD_i: Due Date of part *i*

pc: Penalty cost for unit time delayed

Decision variables:

$$x_{ik} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if part *i* processed in cell *k*
0 otherwise

$$y_{jk} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if machine *j* assigned to cell *k*
0 otherwise

$$Z_{is[r]} = \begin{cases} 1 \\ 0 \end{cases}$$

1 if part *i* assigned to sequence [*r*] in scenario *s*
0 otherwise

$F_{[r]ks}$: The time in which process of part with sequence $[r]$ ends in cell k and scenario s

$FD_{[r]ks}$: Due date of part with sequence $[r]$ in cell k in scenario s

$L_{[r]ks}$: Tardiness of part with sequence $[r]$ in cell k in scenario s

ML_s: Maximum Tardiness occurred in scenario s

D_{iks}: Total processing times of part i needs to be processed in cell k and scenario s

$T_{[r]ks}$: Total processing times of a part with sequence $[r]$ assigned to cell k in scenario s

CF decisions are scenario – independent: they must be made before occurring scenarios and they are made based on their similarities in processing parts and are independent to quantity of processing time. Scheduling decisions are scenario – dependent, thus Z, D, T, FD, L, ML and F variables are indexed by scenario since they must be made after we realize scenario and where the processing time is occurred.

Mathematical Model (Ghezavati, V.R. and Saidi-Mehranad, M., 2010)

$$\text{Minimize } Z = \sum_s pc \times p_s \times ML_s + \sum_k \sum_j \sum_i c_i a_{ij} x_{ik} (1 - y_{jk}) + \sum_k \sum_j \sum_i u_{ij} (1 - a_{ij}) x_{ik} y_{jk} \quad (31)$$

Subject to:

$$\sum_k x_{ik} = 1 \quad \forall i \quad (32)$$

$$\sum_k y_{jk} = 1 \quad \forall j \quad (33)$$

$$\sum_r Z_{is[r]} = 1 \quad \forall i, s \quad (34)$$

$$\sum_i x_{ik} Z_{is,[r+1]} \le \sum_i X_{ik} Z_{is[r]} \quad \forall k, s, r \quad (35)$$

$$D_{iks} = \sum_j a_{ij} t_{ijs} x_{ik} y_{jk} \quad \forall i, k, s \quad (36)$$

$$\sum_i x_{ik} Z_{is[r]} \le 1 \quad \forall r, s, k \quad (37)$$

$$T_{[r]ks} = \sum_i Z_{is[r]} D_{iks} \quad \forall k, s, r \quad (38)$$

$$F_{[r]ks} = \sum_{r=1} \sum_{\alpha=1}^{r} T_{\alpha ks} \quad \forall k, s, r \quad (39)$$

$$FD_{[r]ks} = \sum_i x_{ik} \times Z_{is[r]} \times DD_i \quad \forall k, s, r \quad (40)$$

$$L_{[r]ks} = \max\left\{0, F_{[r]ks} - FD_{[r]ks}\right\} \quad \forall k, s, r \quad (41)$$

$$ML_s = Max\{L_{[r]ks} : k = 1,...,C \text{ } and \text{ } [r] = 1,...,P\} \quad \forall s \quad (42)$$

$$\sum_j y_{jk} \le M_{\max} \quad \forall k \quad (43)$$

$$x_{ik}, y_{jk}, Z_{isr} \sim (0,1) \quad (44)$$

$$D_{iks}, T_{rks}, F_{rks}, FD_{rks} \ge 0 \quad (45)$$

Set constraints (32), (33) and (43) indicate cell formation constraints and set constraints (34), (35), (36), (37), (38), (39), (40), (41) and (42) perform scheduling computations and rational constraints.

Linearization Approaches

In above formulation, since there are both binary and continuous variables where are multiplied to each other, nonlinear terms are appeared in formulation process. Two common types of nonlinear terms are:

Type 1: Pure 0-1 polynomial problem in which n binary variables are multiplied to each other such as $Z = x_1 \times x_2 \times ... \times x_n$.
Type 2: Mixed 0-1 polynomial problems which n binary variables are multiplied to each other and this term is multiplied to a continuous variable such as $Z = x_1 \times x_2 \times ... \times x_n \times Y$.

For linearization type 1 the following method can be applied by introducing some new auxiliary constraints:

$$Z \leq x_i \qquad i = 1, 2, ..., n$$
$$Z \geq \sum_{i-1}^{n} x_i - (n+1)$$

Also, for linearization type 2 in a minimization problem, the following auxiliary constraints will be applied:

P1: Nonlinear problem

$$MinZ = x_1 \times x_2 \times ... \times x_n \times y$$

St:

$$L(X, Y)$$

P2: Linear form

$$Min \quad Z$$

St:

$$Z \geq y - \left|U\right| \times \left(n - \sum_{i=1}^{n} x_i\right)$$
$$Z \geq 0 \qquad L(X, Y)$$

where U is upper bound for continuous variable y and therefore Z will be a continuous variable (Ghezavati and Saidi-Mehrabad 2011).

CONCLUSION

In summary, in this chapter basic principles of uncertainty in a cellular manufacturing system were established. Since CMS problem is affected by tactical decisions such as scheduling, production planning, layout considerations, utilization aspects and many other factors, thus each CMS problem must be aggregated with tactical decisions in order to achieve maximum efficiency. As it is known, tactical decisions are made of many uncertain parameters. Since strategic decisions are influenced by tactical decisions, therefore CMS decisions will be mixed with uncertainty. There are some popular approaches which can analysis uncertain problems such as: Stochastic Optimization, Discrete Planning - Set of Scenario, Continues Optimization, Mean Value model, Mean – Variance Model, Max Probability Optimization, Chance Constrained Programming, Queuing Theory and Markov Chain, and Robust Optimization. This chapter has proposed two sample mathematical models and also one published model [32]. It was assumed that processing routing, inter arrival and service time and also processing time to be uncertain. Stochastic optimization and queuing theory were to resolve uncertainty in formulation process. A complete survey on meta-heuristic methods to solve CMS problems can be found by Ghosh et al (2011). For future directions, the following suggested developments can be applied for researchers and readers:

- Uncertain Processing time optimized by robust approach in continuous or discrete space
- Uncertain capacities optimized by stochastic or robust approach in discrete space

- Uncertain machines' availability optimized by stochastic or queuing theory approaches in continuous or discrete space.
- Aggregating CMS problem with logistics considerations in uncertain environments.
- Aggregating CMS problem with production planning aspects in uncertain environments.
- Aggregating CMS problem with layout considerations in uncertain environments.
- Aggregating CMS problem with scheduling concerns in uncertain environments.

REFERENCES

Andres, C., Lozano, S., & Adenso-Diaz, B. (2007). Disassembly sequence planning in a disassembly cell. *Robotics and Computer-integrated Manufacturing*, *23*(6), 690–695. doi:10.1016/j.rcim.2007.02.012

Aryanezhad, M. B., & Aliabadi, J. (2011). A new approach for cell formation and scheduling with assembly operations and product structure. *International Journal of Industrial Engineering*, *2*, 533–546..doi:10.5267/j.ijiec.2010.06.002

Asgharpour, M. J., & Javadian, N. (2004). Solving a stochastic cellular manufacturing model using genetic algorithm. *International Journal of Engineering, Transactions A. Basics*, *17*(2), 145–156.

Balakrishnan, J., & Cheng, C. H. (2007). Dynamic cellular manufacturing under multi-period planning horizon. *European Journal of Operational Research*, *177*(1), 281–309. doi:10.1016/j.ejor.2005.08.027

Ghezavati, V. R., & Saidi-Mehrabad, M. (2011). An efficient hybrid self-learning method for stochastic cellular manufacturing problem: A queuing-based analysis. *Expert Systems with Applications*, *38*, 1326–1335. doi:10.1016/j.eswa.2010.07.012

Ghezavati, V. R., & Saidi-Mehranad, M. (2010). Designing integrated cellular manufacturing systems with scheduling considering stochastic processing time. *International Journal of Advanced Manufacturing Technology*, *48*(5-8), 701–717. doi:10.1007/s00170-009-2322-2

Ghezavati, V. R., & Saidi-Mehranad, M. (2011). An efficient linearization technique for mixed 0-1 polynomial problems. *Journal of Computational and Applied Mathematics*, *235*(6), 1730–1738. doi:10.1016/j.cam.2010.08.009

Ghosh, T., Sengupta, S., Chattopadhyay, M., & Dan, P. K. (2011). Meta-heuristics in cellular manufacturing: A state-of-the-art review. *International Journal of Industrial Engineering Computations*, *2*(1), 87–122. doi:10.5267/j.ijiec.2010.04.005

Heragu, S. (1997a). *Facilities design* (p. 316). Boston, MA: PWS Publishing Company.

Heragu, S. (1997b). *Facilities design,* (p. 345). Boston, MA: PWS publishing company

Hillier, F. S., & Lieberman, G. J. (1995). *Introduction to operation research* (6th ed.). New York, NY: McGraw-Hill.

Hosseini, M. M. (2000). An inspection model with minimal and major maintenance for a system with deterioration and Poisson failures. *IEEE Transactions on Reliability*, *49*(1), 88–98. doi:10.1109/24.855541

Hurley, S. F., & Clay Whybark, D. (1999). Inventory and capacity trade-off in a manufacturing cell. *International Journal of Production Economics*, *59*(1), 203–212. doi:10.1016/S0925-5273(98)00101-7

Kuroda, M., & Tomita, T. (2005). Robust design of a cellular-line production system with unreliable facilities. *Computers & Industrial Engineering*, *48*(3), 537–551. doi:10.1016/j.cie.2004.03.004

Mahdavi, I., Javadi, B., Fallah-Alipour, K., & Slomp, J. (2007). Designing a new mathematical model for cellular manufacturing system based on cell utilization. *Applied Mathematics and Computation*, *190*, 662–670. doi:10.1016/j.amc.2007.01.060

Mobasheri, F., Orren, L. H., & Sioshansi, F. P. (1989). Scenario planning at southern California Edison. *Interfaces*, *19*(5), 31–44. doi:10.1287/inte.19.5.31

Mulvey, J. M. (10996). Generating scenarios for the Towers Perrin investment system. *Interfaces*, *26*(2), 1–15. doi:10.1287/inte.26.2.1

Papaioannou, G., & Wilson, J. M. (2009). Fuzzy extensions to integer programming of cell formation problem in machine scheduling. *Annals of Operations Research*, *166*(1), 1–19. doi:10.1007/s10479-008-0423-1

Ravichandran, K. S., & Chandra Sekhara Rao, K. (2001). A new approach to fuzzy part family formation in cellular manufacturing system. *International Journal of Advanced Manufacturing Technology*, *18*(8), 591–597. doi:10.1007/s001700170036

Rosenhead, J., Elton, M., & Gupta, S. K. (1972). Robustness and optimality as criteria for strategic decisions. *Operational Research Quarterly*, *23*(4), 413–431.

Safaei, N., Saidi-Mehrabad, M., Tavakkoli-Moghaddam, R., & Sassani, F. (2008). A fuzzy programming approach for cell formation problem with dynamic & uncertain conditions. *Fuzzy Sets and Systems*, *159*(2), 215–236. doi:10.1016/j.fss.2007.06.014

Shanker, R., & Vrat, P. (1998). Post design modeling for cellular manufacturing system with cost uncertainty. *International Journal of Production Economics*, *55*(1), 97–109. doi:10.1016/S0925-5273(98)00043-7

Siemiatkowski, M., & Przybylski, W. (2007). Modeling and simulation analysis of process alternative in cellular manufacturing of axially symmetric parts. *International Journal of Advanced Manufacturing Technology*, *32*(5-6), 516–530. doi:10.1007/s00170-005-0366-5

Snyder, L. V. (2006). Facility location under uncertainty: A review. *IIE Transactions*, *38*, 537–554. doi:10.1080/07408170500216480

Solimanpur, M., Vrat, P., & Shankar, R. (2004). A heuristic to optimize makespan of cell scheduling problem. *International Journal of Production Economics*, *88*, 231–241. doi:10.1016/S0925-5273(03)00196-8

Song, S.-J., & Hitomi, K. (1996). Determining the planning horizon and group part family for flexible cellular manufacturing. *Production Planning and Control*, *7*(6), 585–593. doi:10.1080/09537289608930392

Sun, Y.-L., & Yih, Y. (1996). An intelligent controller for manufacturing cell. *International Journal of Production Research*, *34*(8), 2353–2373. doi:10.1080/00207549608905029

Szwarc, D., Rajamani, D., & Bector, C. R. (1997). Cell formation considering fuzzy demand and machine capacity. *International Journal of Advanced Manufacturing Technology*, *13*(2), 134–147. doi:10.1007/BF01225760

Tsai, C. C., Chu, C. H., & Barta, T. (1997). Analysis and modeling of a manufacturing cell formation problem with fuzzy integer programming. *IIE Transactions*, *29*(7), 533–547. doi:10.1080/07408179708966364

Venkataramanaiah, S. (2007). Scheduling in cellular manufacturing systems: An heuristic approach. *International Journal of Production Research*, *1*, 1–21.

Viswanadham, N., & Narahari, Y. (1992). *Performance modeling of automated manufacturing systems*. Englewood Cliffs, NJ: Prentice Hall.

Wu, X. D., Chu, C. H., Wang, Y. F., & Yan, W. L. (2006). Concurrent design of cellular manufacturing systems: A genetic algorithm approach. *International Journal of Production Research*, *44*(6), 1217–1241. doi:10.1080/00207540500338252

Yang, J., & Deane, R. H. (1993). Setup time reduction and competitive advantage in a closed manufacturing cell. *European Journal of Operational Research*, *69*(3), 413–423. doi:10.1016/0377-2217(93)90025-I

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 298-316, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 32
Multi-Modal Assembly-Support System for Cellular Manufacturing

Feng Duan
Nankai University, China

Jeffrey Too Chuan Tan
The University of Tokyo, Japan

Ryu Kato
The University of Electro-Communications, Japan

Chi Zhu
Maebashi Institute of Technology, Japan

Tamio Arai
The University of Tokyo, Japan

ABSTRACT

Cellular manufacturing meets the diversified production and quantity requirements flexibly. However, its efficiency mainly depends on the operators' working performance. In order to improve its efficiency, an effective assembly-support system should be developed to assist operators during the assembly process. In this chapter, a multi-modal assembly-support system (MASS) was proposed, which aims to support operators from both information and physical aspects. To protect operators in MASS system, five main safety designs as both hardware and control levels were also discussed. With the information and physical support from the MASS system, the assembly complexity and burden to the assembly operators are reduced. To evaluate the effect of MASS, a group of operators were required to execute a cable harness task. From the experimental results, it can be concluded that by using this system, the operators' assembly performance is improved and their mental work load is reduced. Consequently the efficiency of the cellular manufacturing is improved.

DOI: 10.4018/978-1-4666-1945-6.ch032

INTRODUCTION

Traditionally, when the mass production was major in industry production, various assembly systems had been designed as automated manufacturing lines, which are aimed to produce a single specific product without much flexibility. Nowadays, the tastes of consumers change from time to time; therefore, traditional automated manufacturing lines cannot meet the flexibility and efficiency at the same time. To solve this problem, cellular manufacturing system, also called cell production system, has been introduced. In this system, an operator manually assembles each product from start to finish (Isa & Tsuru, 2002; Wemmerlov & Johnson, 1997). The operator enables a cellular manufacturing system to meet the diversified production and quantity requirements flexibly. However, due to the negative growth of the population in Japan, it will become difficult to maintain the cellular manufacturing system with enough skilled operators in the near future. How to improve the assembly performance of the operators and how to reduce their assembly burden are two important factors, which limit the efficiency of the cellular manufacturing system.

Without an effective supporting system, it is difficult to maintain the cellular manufacturing system in Japan. Taking the advantages of the operators and robots, but avoiding their disadvantages at the same time, a new cellular manufacturing system was proposed, namely, the human-robot collaboration assembly system (Duan, 2008). In this human-robot collaboration assembly system, the operators are only required to execute the complicated and flexible assembly tasks that need human assembly skills; while the robots are employed to execute the monotonous and repeated tasks, such as the repetitions of parts feeding during assembly process (Arai, 2009). To make this system has the applicability to assemble a variety of products in different manufacturing circumstances, the following assembly sequence is assumed: each assembly part is collected from the tray shelf by manipulators; all the parts are automatically fed to the operator on a tray as a kit of parts; an operator grasps the individual part respectively and assembles it to form a final product; the assembled product is transferred out to the next station, and so on.

In the following part, a multi-modal assembly-support system (MASS) is introduced, which aims to support an assembly operator in a cellular manufacturing system from both information side and physical side while satisfying the actual manufacturing requirements. MASS system utilizes robots to support the operator and several information devices to monitor and guide the operator during the assembly process. Since it is a human-robot collaboration assembly system, safety strategy must be designed to protect the operator with a reasonable cost benefit balance in the real production line.

The remainder of the chapter is organized as follows: Firstly, the background information and the related studies are introduced. Then, the entire MASS system and its subsystems are briefly described. After that, a description of two manipulators and a mobile base are introduced in physical support part, which are used to feed assembly parts to the assembly operator. Assembly information support part contains a discussion of a multimedia-based assembly table and corresponding devices. Safety standard and safety design are presented in safety strategy part. Taking a cable harness task as an example, the effect of MASS system was evaluated. Finally, the conclusion and the future work are given.

PREVIOUS RELATED STUDIES

To improve the efficiency of the cellular manufacturing system, various cellular manufacturing systems have been designed to improve the assembly performance of the operators and reduce their assembly burden.

Seki (2003) invented a production cell called "Digital Yatai" which monitors the assembly progress and presents information about the next assembly process. Using a semi-transparent head mount display, Reinhart (2003) developed an augmented reality (AR) system to supply information to the operator. These studies support the operator from information aspect. To reduce the operator's physical burden and improve the assembly precision, Hayakawa (1998) employed a manipulator to grasp the assembly parts during the assembly process. This improved the assembly cell in physical support aspect. Sugi (2005) aimed to support the operators from both information side and physical side, and developed an attentive workbench (AWB) system. In this system, a projector was employed to provide assembly information to the operator; a camera was used to detect the direction of an operator's pointing finger; and several self-moving trays were used to deliver parts to the operator. Although AWB achieved its goal of supporting operators from both information aspect and physical aspect, the direct supporting devices are just a projector and several self-moving trays, which are general purpose instruments that cannot meet the actual manufacturing requirements.

In the coming aging society, it will be impossible to maintain the working efficiency if everything is done manually by the operator in the current cellular manufacturing system. In order to increase working efficiency, many researchers have used robot technologies to provide supports to the operator (Kosuge, 1994; Bauer, 2008; Oborski, 2004). According to these studies, human-robot collaboration has potential advantages to improve the operator's working efficiency. However, before implementing this proposal, the most fundamental issue will be the safety strategy, which allows the operators and the robots to execute the collaboration work in their close proximity.

Human-robot collaboration has been studied in many aspects but has not been utilized in the real manufacturing systems. This is mainly because safety codes on industrial robots (ISO 12100, ISO 10218-1, 2006) prohibit the coexistence of an operator in the same space of a robot. According to the current industrial standards and regulations, in a human-robot collaboration system, a physical barrier must be installed to separate the operator and the assisting robot. Under this condition, the greatest limitation is that the close range assisting collaboration is impossible. Based on the definition of Helms (2002), there are four types of human-robot collaboration: *Independent Operation*, *Synchronized Cooperation*, *Simultaneous Cooperation*, and *Assisted Cooperation*. The assisted cooperation is the closest type of collaboration, which involves the same work piece being processed by the operator and the robot together. In this kind of human-robot collaboration, the operator is working close to the working envelope of the assisting robot without physical separation, so that both of them can work on the same work piece in the same process. The most distinguished concept of this study is that the assisting robot in this work is active and is able to work independently as robot manipulator. The advantage of this collaboration is to provide a human-like assistance to the operator, which is similar with the cooperation between two operators. This kind of assistance can improve the working efficiency by automating portion of the work and enable the operator to focus only on the other portion of work which requires human skill and flexibility. However, since the active robot is involved, this kind of collaboration is extremely dangerous and any mistake can be fatal (Beauchamp & Stobbe, 1995).

The challenge of this research work is to design an effective assembly supporting system, which can support the operator in both physical and information aspects. During the assembly process, employing of the assisting robot is an effective method to reduce the operator's assembly burden while improving the working efficiency. This involves the safety issue in this kind of close range active human-robot collaboration. However, there are no industrial safety standards and regulations. Besides the design of the assembly supporting system, the scope of this work also covers both safety design study and development of prototype production cell in cellular manufacturing.

MULTI-MODAL ASSEMBLY-SUPPORT SYSTEM

Structure of the Entire System

Following the fundamental idea that robots and operators share the assembly tasks can maximize their corresponding advantages, the MASS system was designed and its subsystems are shown in Figure 1 as structure view and in Figure 2 as system configuration.

The entire MASS system is divided into physical support part and assembly information support part, as shown in Figure 1.

1. **Physical Supporting Part**: The physical supporting part is aimed to support operators from physical aspect, and it is composed of two manipulators with six degrees of freedom and a mobile base, which have two functions: one is to deliver assembly parts from a tray shelf to an assembly table; and the other is to grasp the assembly parts and prevent any wobbling during the assembly process.
2. **Information Supporting Part**: The assembly information supporting part is designed to aid operators in assembly information aspect. An LCD TV, a speaker, and a laser pointer are employed to provide assembly information to guide the operator.

Figure 1. Structure of the entire MASS system

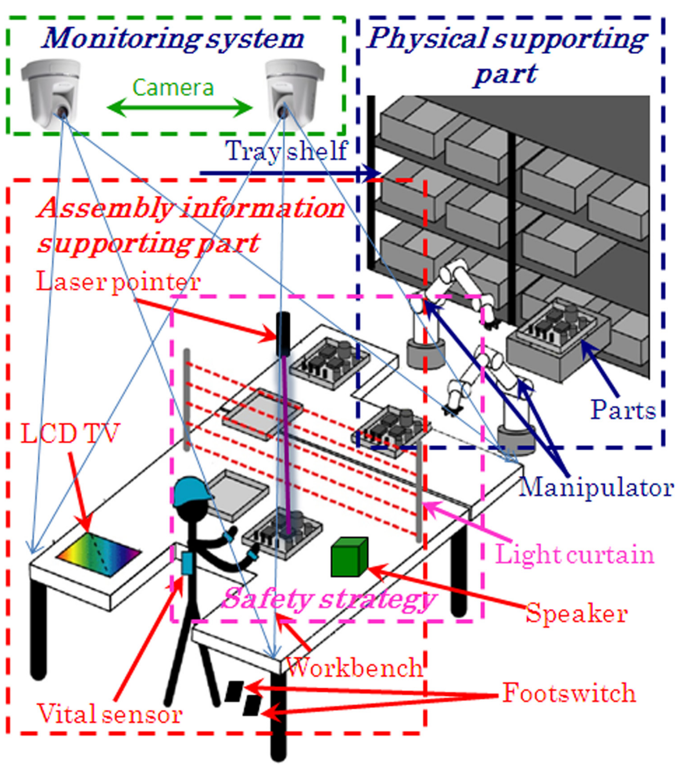
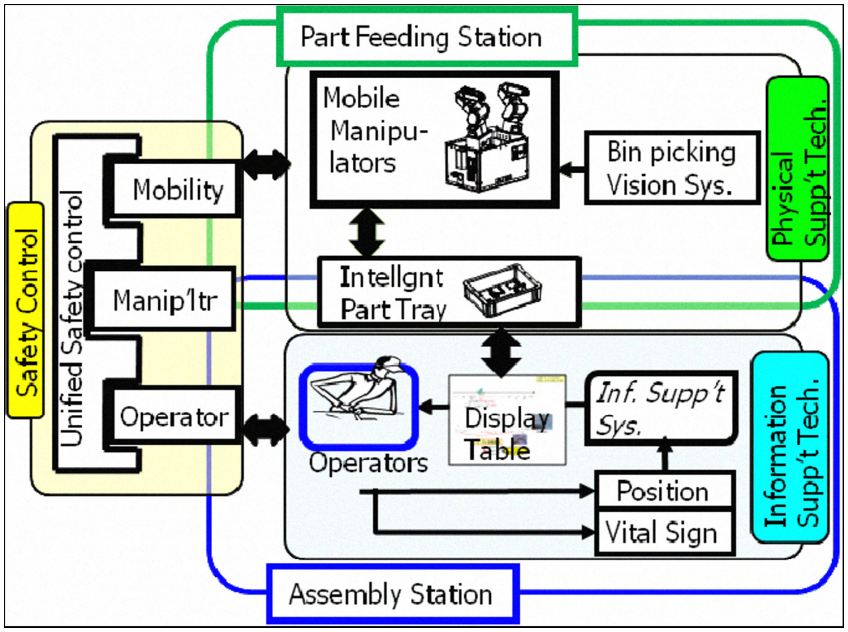


Figure 2. Configuration of MASS system



3. **Safety Control Part**: To guarantee the operator's safety during the assembly process, vital sensors are used to monitor the operator's physical conditions during the assembly process, and a serial of safety strategies is used to protect the operator from injury by the manipulators. It controls the collaboration between a robot and an operator (also referred to Figure 2).

In the developed MASS system, there are two stations connected through an intelligent part tray as shown in Figure 2, on which all the necessary parts are fed into the assembly station and the assembled products are transferred out through a shipment from the assembly station.

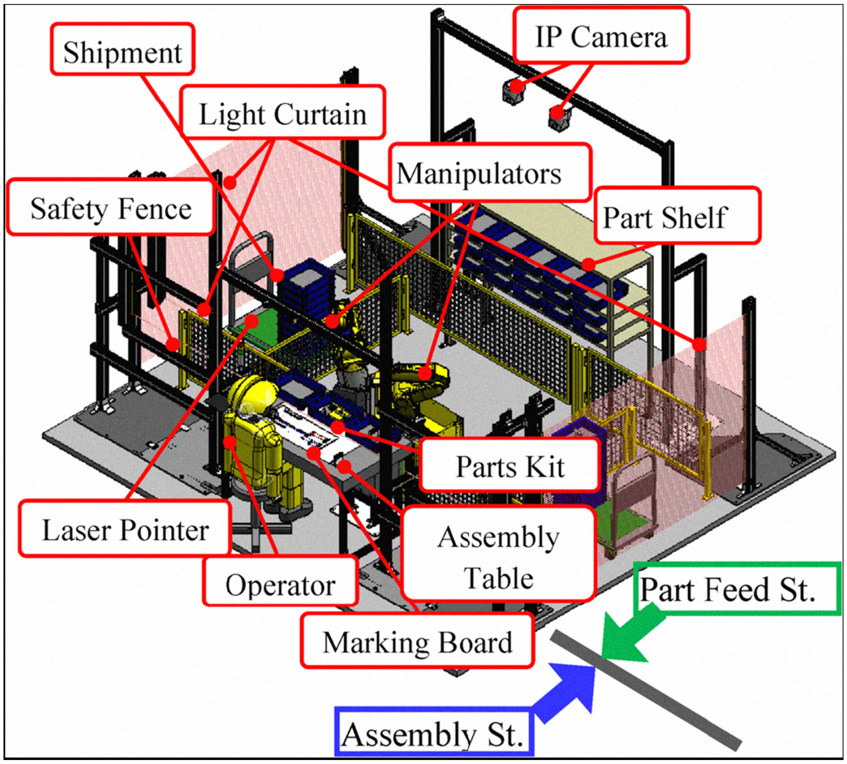
1. **Part Feeding Station**: Only robots work here. It is mainly in charge of part handling, such as bin picking, part feeding, kitting and part transferring.
2. **Assembly Station**: An operator executes the assembly tasks with some aid of the robots. Supporting **information** from the MASS system is implemented to accelerate the operator assembly efficiency.

Figure 1 illustrates the setup of the MASS system, in which an operator assembles a product on the workbench in the area of assembly station. The operator is supported with the assembly information and with physical holding of parts for assembly. In this study, the sample product to assemble is a cable harness with several connectors and faster plates. Even experienced operators maybe spend about 15 minutes finishing this assembly task.

Simulator of the Entire System

To reduce the design period, in this study, a simulator of the entire system was developed based on ROBOGUIDE (FANUC ROBOGUIDE) and OpenGL (Neider, 1993), as shown in Figure 3. This simulator can not only reproduce the actual motion of the manipulators but also predict collisions in the work space.

Figure 3. Simulator of the entire MASS system



Since MASS system is a human-robot cooperation assembly system, considering the operator's safety, the distance between the manipulators and the operator should be optimized to prevent the collisions between them. Furthermore, the moving trajectories of the manipulators should also be optimized to prevent the collisions between themselves. In order to accelerate the development period, all of the optimization assignments are done in this simulator first, and then evaluated in the actual MASS system. With the aid of this simulator, the distance between manipulators and the operator can be adjusted easily, and the moving trajectories of the manipulators' end points can also be reproduced conveniently during the manipulators' moving process. Therefore, based on the simulation results, the actual system could be conveniently constructed.

Physical Support

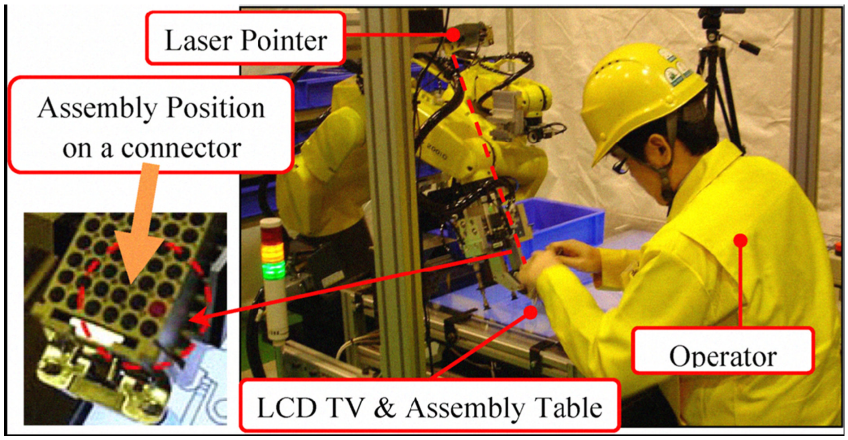
To increase the physical support provided by the MASS system, two manipulators with six degrees of freedom are installed on a mobile base and used to deliver assembly parts to the operator, as shown in Figures 1-4. A CCD camera with an LED light is equipped to each manipulator respectively for recognition of picking target from a part bin in scramble.

The manipulators are utilized in part feeding station to

1. Draw a part bin from part shelves;
2. Pick a part from the bin one by one;
3. Kit parts onto a tray;
4. Check visually the parts in a tray.

The parts are efficiently fed by the manipulators, because one manipulator hangs a bin up and

Figure 4. Assembly operations with the aid of manipulators



the other one grasps a part out like an operator does. Since the bin picking system by manipulators can work 24 hours a day, it enables high productivity. The base carries a few trays and moves to the assembly station, where the base docks in the electric charge connector. In the assembly station, an operator continuously assembles parts one by one, which are transferred by one of the mobile twin manipulators. To increase the precision of assembly and reduce the operator's burden, one manipulator can grasp an assembly part to prevent wobbling during assembly, and the operator executes the assembly task on the basis of the manipulator's assistance, as shown in Figure 4.

Obviously, the assistant manipulators move near to the operator during the assembly process. To achieve this collaboration, the manipulators have to penetrate the operator's area. Since the penetration is prohibited by the regulations of the industrial robots (ISO 12100), a new countermeasure must be developed. After finishing an assembly step, the operator pushes a footswitch to send a control command to the manipulators, and the manipulators provide the next assembly part to the operator and the assembly information of the next assembly step is given. Without this control command, the manipulators cannot move to the next step. Furthermore, the operator can stop the manipulators with an emergency button when an accident occurs. These strategies enable the manipulators to support human operators in physical aspect effectively and safely.

Assembly Information Support

Previous studies, Szeauch as Digital Yatai (Seki, 2003), have already testified that providing assembly information to the operator during his assembly process can not only improve his assembly efficiency, but also reduce his assembly errors. Taking the advantages of the previous studies, and also considering the characteristics of human cognition, an assembly information supporting system is designed to guide operators by means of indicating the next assembly sequence and/or an appropriate way of operation.

The developed system has three major advantages:

1. Each assembly sequence is instructed step by step;
2. Considering the characteristics of human cognition, the assembly information can be provided as easily understandable formats for humans, including text, voice, movie, animation and flashing emphasis marks;

3. The assembly information can be selected and provided to the operator according to his assembly skill level.

The total software system of MASS system in Figure 5 has been developed. It consists of three subsystems as

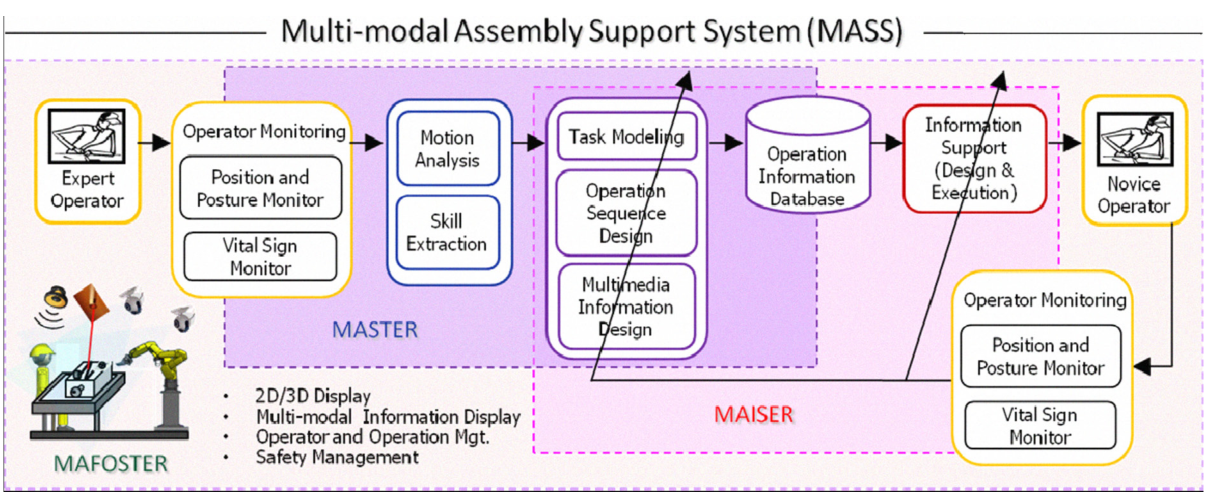
1. Multi-modal Assembly Skill TransfER (MASTER);
2. Multi-modal Assembly Information SupportER (MAISER);
3. Multi-modal Assembly FOSTER (MAFORSTER).

MASS is designed to extract the skill information from skilled operators by MASTER and to transfer it to novice operators by MAISER as illustrated in Figure 5. Here, a human assembly skill model was proposed (Duan, 2009), which extracts and transfers human assembly skills as the cognition skill part and the motor skill part. In the cognition skill part, depending on questionnaire, MASTER obtains the different cognition skills between the skilled operators and the novice operators. In the motor skill part, MASTER mainly utilizes motion capture system to obtain the different motor skills between the skilled operators and the novice operators, especially in the assembly pose aspect and the assembly motion aspect (Duan, Tan, Kato, & Arai, 2009). MAISER provides understandable instructions to novice operators by displaying multi-modal information about assembly operations. MAFOSTER controls interface devices to organize comfortable environment for operators to execute the assembly task like a foster does. MAISER works mainly off-line at a data-preparation phase, and watches on-line the state of an operator to avoid bad motion and dangerous states (Duan, Tan, Kato, & Arai, 2009). MAISER takes the role of an instruction phase.

Interface devices are installed as shown in Figure1 and Figure 4 again:

1. **LCD TV:** The horizontal assembly table with built-in 37 inch LCD TV as shown in Figure 4 may be the first application for assembly. Since it enables operators to read the instructions without moving his/her gaze in different direction, assembly errors can be decreased. The entire assembly scheme is divided into several simple assembly steps, and the corresponding assembly information is written in PowerPoint slides (Zhang, 2008). During the assembly process, these

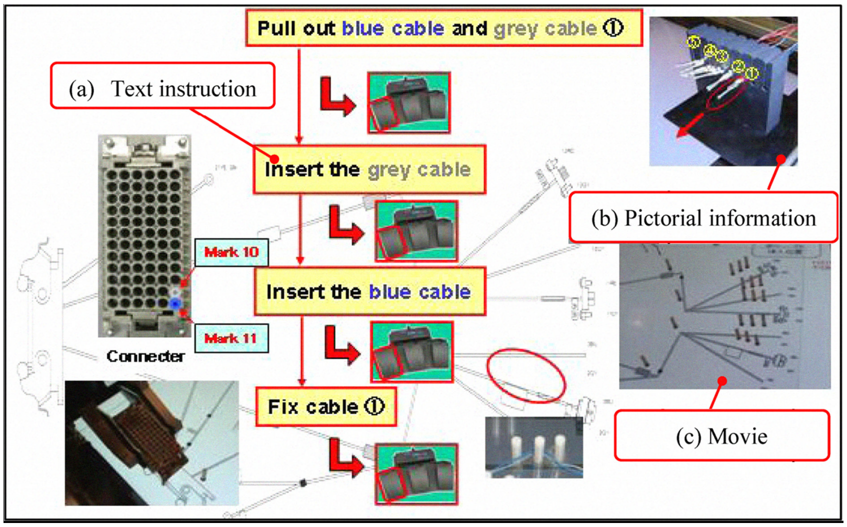
Figure 5. Software system of MASS system



PowerPoint slides are inputted into the LCD TV and switched by footswitch.

2. **Laser Pointer:** Showing the assembly position to the operator is an effective way to reduce assembly mistakes. To this end, a Laser pointer, which is fixed on the environment, is projected onto a task to indicate the accurate position of assembly as shown in the left photo of Figure 4. The position can change by the motion of the manipulator. The operator can insert a wire into the instructed assembly position with the aid of the Laser spot.
3. **Audio Speakers:** To easily permit the operator to understand the assembly information, a speaker and a wireless Bluetooth earphone are used to assist the operator with voice information.
4. **Footswitch:** During the assembly process, it is difficult for the operator to switch the PowerPoint slides with his hands. Therefore, a footswitch is used, as shown in Figure 1. There are two kinds of footswitches: footswitch A has three buttons, and footswitch B has one button. Just stepping the different buttons on footswitch A, the operator can move the PowerPoint slides forward or backward. Stepping the button on footswitch B, the operator controls the manipulators to supply the necessary assembly parts to the operator, or makes manipulators change the position and orientation of the assembly part during the assembly process.
5. **Assembly Information:** The assembly support information is provided to the operators to improve the productivity by means of good understanding in assembly tasks and of skill transfer with audio-visual aids. As the software structure for the assembly task description is not discussed in this study, please refer to our papers (Duan, 2008; Tan, 2008). Applying Hierarchical Task Analysis (HTA) one assembly task is divided into several simpler assembly steps, whose corresponding information is stored in multimedia. Then appropriate level of information is displayed on LCD panel as shown in Figure 6. In each PowerPoint slide, the assembly parts and assembly tools are illustrated with pictures. The assembly positions are noted with color marks. Following the assembly flow chart, videos showing the assembly motions of the experienced operators will appear to guide the novices to execute the assembly tasks. To facilitate the operator's understanding of the assembly process, the colors of the words in the slides are the same as the actual colors used for the assembly parts. For example, there are "blue cable" and "grey cable" in Figure 6. In each slide, several design principals of data presentation are introduced such as multimedia principle, coherence principle and spatial contiguity principle (Mayer, 2001). In Figure 6, three types of information are displayed as (a) text instruction, (b) pictorial information, (c) movie, and the sequence of assembly is also illustrated. During the assembly process, the PowerPoint slides are output to an LCD TV and switched by the operator's foot with footswitch during the assembly process.
6. **Assembly Information Database:** In this multimedia based assembly supporting system, the assembly information is classified into paper, audio, and video files. The assembly guidance is concisely written in paper files. Guidance of each assembly step is recorded in audio files. After the standard motions of the experienced operators are recorded and analyzed into primitive assembly motions, they are saved into video files. Tan (2008) set up an assembly information database to preserve all of these assembly information files and provide them to the operator depending on the situation. This database contains training data and assembly data: training data are designed for novices, and the assembly information files contain

Figure 6. Multimedia based assembly supporting information



assembly details. Assembly data are used to assist experienced operators by indicating the assembly sequence but not assembly details. As a consequence, this system may promote both novice and experienced operators to enter the workforce.

All the operators who used the assembly table with LCD evaluated positively that the instruction on LCD can be read easily and understood smoothly.

Safety Strategy

MASS system is a kind of human-robot cooperation system. Although employing the assistant robots to support the operator can increase the assembly efficiency and reduce the assembly burden, this collaboration can be extremely dangerous because the active robot is involved and any mistake can be fatal. To protect the operator during the assembly process, several safety designs are proposed and developed in this manufacturing system, which cover both hardware and software to achieve good robot-human collaboration. Fundamental concepts are:

1. Risk assessment by ISO regulation;
2. Area division by safety light curtains as illustrated in Figure 7;
3. Speed/Force limiter by serve controller;
4. Collision protector by physical devices;
5. Collision detector by IP cameras;
6. Inherent safety theory.

Risk Assessment by ISO Regulation

Since no direct industrial safety standards and regulations that govern this type of close range active human-robot collaboration, the safety design in this work is formulated by collective reference to related safety standards and regulations to verify component systems' safety first (non-collaboration safety) and then assess system safety as a whole (collaboration safety). Table 1 summarizes the referred industrial safety standards and regulations in mobile robot manipulators system development and total system development.

This chapter mainly focuses on the discussion on human-robot collaboration safety; therefore the non-collaboration safety of component systems is omitted. However, it is important to bear in mind that the following safety designs for collaboration are built in accordance with the referred

Table 1. Related safety standards and regulations

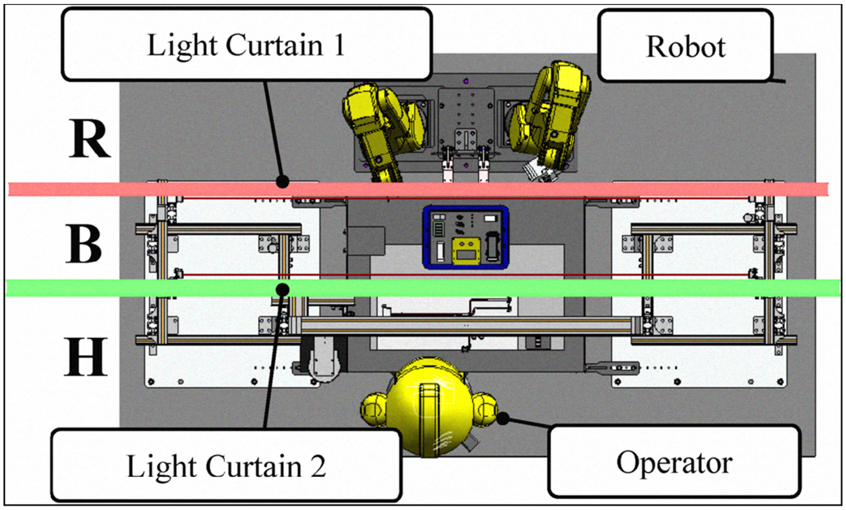
Standards and Regulations	Descriptions
Related to mobile robot manipulators system development	
IEC 60364-4-41 (JIS C0364-4-41)	Low-voltage electrical installations – Part 4-41: Protection for safety – Protection against electric shock
IEC 60364-7-717	Electrical installations of buildings – Part 7-717: Requirements for special installations or locations – Mobile or transportable units
IEC 61140 (JIS C0365)	Protection against electric shock – Common aspects for installation and equipment
BS EN 1175-1	Safety of industrial trucks – Electrical requirements – Part 1: General requirements for battery powered trucks
ISO 10218-1 (JIS B8433-1)	Robots for industrial environments – Safety requirements – Part 1: Robot
Related to total system development	
ISO 12100-1 (JIS B9700-1)	Safety of machinery – Basic concepts, general principles for design – Part 1: Basic terminology, methodology
ISO 12100-2 (JIS B9700-2)	Safety of machinery – Basic concepts, general principles for design – Part 2: Technical principles
ISO 14121-1 (JIS B9702)	Safety of machinery – Risk assessment – Part 1: Principles
ISO 14121-2	Safety of machinery – Risk assessment – Part 2: Practical guidance and examples of methods
ISO 13849-1 (JIS B9705-1)	Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design
BS EN 954-1	Safety of machinery. Safety related parts of control systems. General principles for design
ANSI/RIA R15.06	Industrial Robots and Robot Systems - Safety Requirements
ISO 13852 (JIS B9707)	Safety of machinery – Safety distances to prevent danger zones being reached by the upper limbs
ISO 14119 (JIS B9710)	Safety of machinery – Interlocking devices associated with guards – Principles for design and selection
ISO 13854 (JIS B9711)	Safety of machinery – Minimum gaps to avoid crushing of parts of the human body
ISO 14118 (JIS B9714)	Safety of machinery – Prevention of unexpected start-up
ISO 13855 (JIS B9715)	Safety of machinery – Positioning of protective equipment with respect to the approach speeds of parts of the human body
ISO 14120 (JIS B9716)	Safety of machinery – Guards – General requirements for the design and construction of fixed and movable guards

standards and regulations in the component level. EU standard permits the collaboration of robots with the operator when the total output of robots is less than 150 (N) at the tip of the end-effecter. Japanese standard defines that each actuator has the power less than 80 (W). The collaboration safety design is presented in hardware design and control design in the following.

Area Division by Safety Light Curtains

The software systems in robot controller and other computers are prepared as Dual Check Safety (DCS), which checks speed and position data of motors with two independent CPUs in the robot controller. In risk assessment, we listed up to 168 risks and take its countermeasure respectively

Figure 7. Three robot working zones for safety



so as to satisfy the required performance level. Whatsoever definition industrial robots are, it is strongly prohibited that robots exist with the operator in the same space. Thus a cage is required to separate the operator from the robots. For the area division, the whole cell in Figure 7 is divided into human area (**H**), robot area (**R**), and buffer area (**B**) by safety fences, photoelectric sensors and light curtains in order to obtain safe working areas and to monitor border crossing for safety. Robots are allowed to operate in high speed motion in area **R** but low speed movement in area **B**. In area **H**, the strong restrictions are applied to robot motions. When the manipulators move too close to the operators and cross the light curtain 2, the power of the manipulator is cut down by the light curtain. Consequently, the manipulators stop.

Speed/Force Limiter by Serve Controller

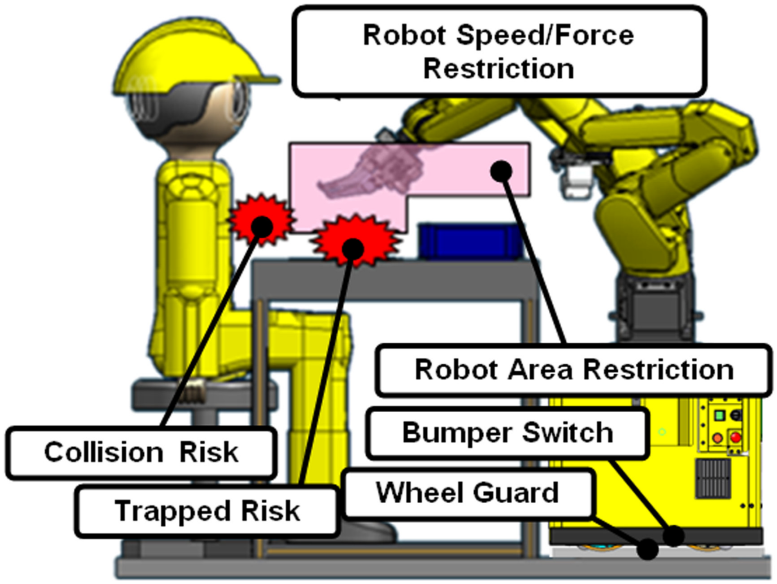
As shown in Figure 8, by the servo controller, the speed of the mobile manipulators is limited, and the force/torque at the end-effecter is also limited by software. The controller also has a function of abnormal force limiter in case of unexpected collision of the manipulator against the environment. Based on the recommendation from safety standards and risk assessment, during collaboration process, the speed of the mobile manipulators is limited to below 150 (mm/s) and the working area of the robot is restricted within the pink region in Figure 8. The minimum distance between the robot gripper and surface of the workbench is 120 (mm) according to ISO 13854.

Collision Protectors by Physical Devices

During the assembly process, several collision protectors by physical devices have been designed for accident avoidance and the protection of the operator.

1. **Mobile Base:** To prevent the operator from being hurt by the manipulators, the localization accuracy of the mobile base should be maintained. With vision system to detect marks on the floor, the system has a localization accuracy of 5 (mm) and 0.1°. The base is equipped with bumper switch for object collision detection and wheel guard to prevent foreign object being tangled with the wheels, as illustrated in Figure 8.

Figure 8. Robot speed, force, and area restrictions



2. **Footswitch:** In the MASS system, the twin mobile manipulators are used to assist the operator to execute the assembly tasks during the assembly process. Without a safety strategy, the operator could be injured by the manipulators. An effective working sequence is one of the effective ways to reduce the probability of collision between an operator and a manipulator. The manipulators are prevented from moving in the direction of the operator as he performs an assembly task. The probability for collisions is reduced with the introduction of the working sequence. To realize the proposed working sequence, a footswitch is used to control the manipulators, as illustrated in Figure 9. When the operator finishes an assembly step, he steps on footswitch, which signals the manipulators to provide the assembly parts to the operator for the next step.
3. **Emergency Button:** When an accident occurs, the operator can just push the emergency button on the right-hand side of the assembly workbench to stop the entire system, as shown in Figure 9. After any problem has been solved, the operator pushes the reset button to restart the assembly process.
4. **Safe Bar:** In addition, steel safe bar is installed in front of the assembly workbench (referred to Figure 9). If other strategies failed to stop the manipulator to collide the operator, this safe bar can protect the operator.

Collision Detector by IP Cameras

The developed system installs a robot with higher ability than both EU and Japan Standard. Even though various countermeasures are introduced, the risk assessment shows residual risks. For the intelligent compensation of safety, two IP cameras are utilized to monitor the operator's safety (referred to Figure 10); that is, the cameras track the color marks on the head and shoulders of the operator to measure the body posture and position to estimate the human operation conditions (Duan, 2009). The vision monitoring system has positioning accuracy of 30 (mm) and process delay of 0.6 (s).

Figure 9. Collision protectors by physical devices

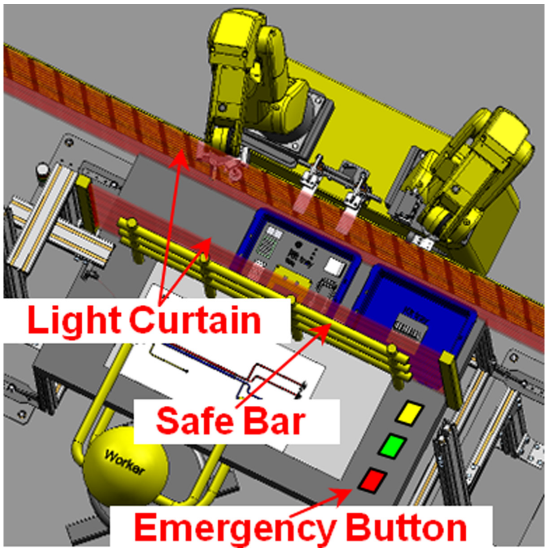


Figure 10. Operator safety monitoring system

Inherent Safety Theory

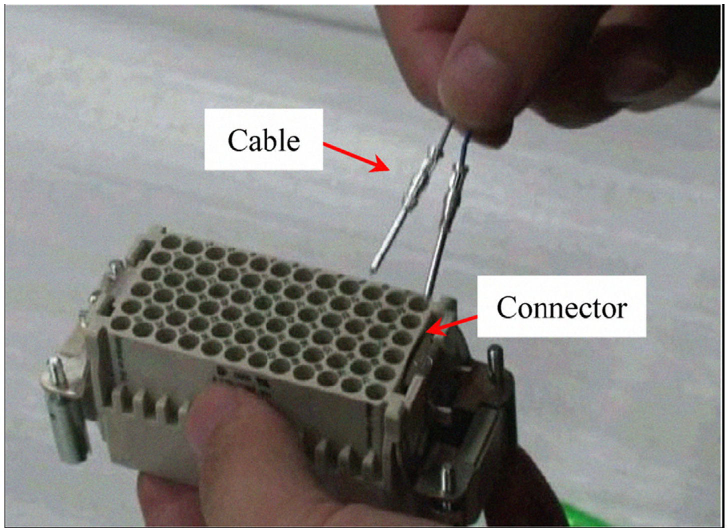
Although several safety strategies are adopted, there is no guarantee that a collision between a manipulator and an operator will never occur. Therefore, the manipulators should be ameliorated according to inherent safety theory (Ikeda & Saito, 2005) to reduce the injury of the operator. The sharp edges of the manipulators are softened into obtuse-angled brims. The force and speed of the manipulators are reduced as much as possible while still meeting the assembly requirement. In addition, the overall mobile robot manipulators system is built with low center of gravity design to prevent tipping.

Evaluation of MASS System

To evaluate the effect of the MASS system, a group of operators were required to execute an assembly task of cable harness, as illustrated in Figure 11. In this task, operators must insert correct cables into corresponding correct holes in the connector. After that, following the cable routes, operators must fix the cables to the jigs on the assembly table. The operators executed the cable harness task in two cases: (1) all of the assembly information, including the cable type, the position of the hole in the connector and the assembly step, was only provided by the assembly manual (Exp I); (2) operators executed the cable harness task under the support of MASS (Exp II). Two parameters were measured in the experiments: assembly time and assembly error. The assembly time is compared between conventional manual assembly setup (Exp I) and the new setup (Exp II). Five novice operators and five expert ones performed three assembly trails respectively for both the setups. From Figure 12, it is proved that the overall performance is better (shorter in assembly time) in the new setup (Exp II). Novices and experts show almost the same assembly time from the first trial in the case of the new setup as the dotted lines.

It means that the assembly can be executed at the minimum time even by the unskilled operators. Comparing to the assembly time of the conventional setup (Exp I), the novice operators need only 50% of the time in the MASS system (Exp II), which indicates double productivity. Note that

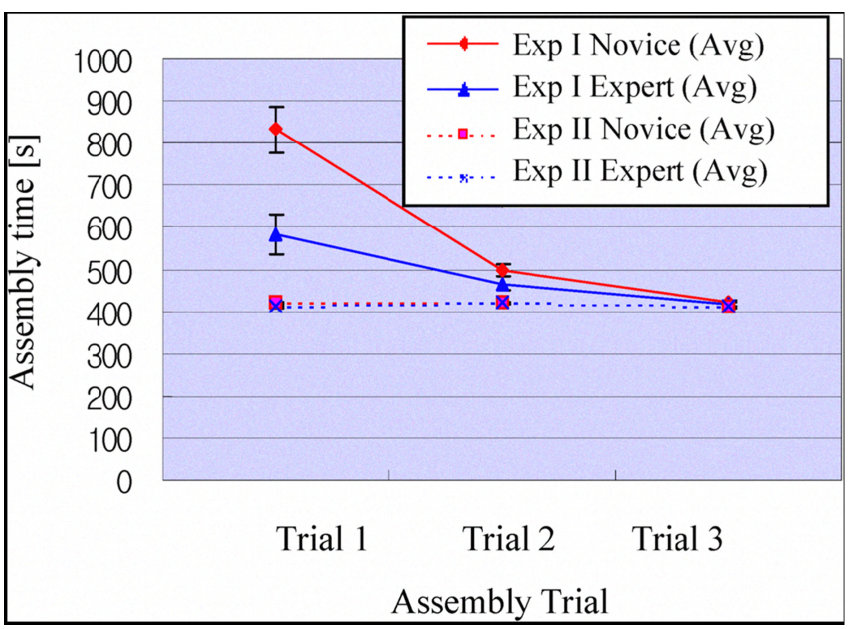
Figure 11. Cable harness task



the assembly time at the third trial converges to the minimum by all the cases. This implies that the assembly operation is easy to achieve and the human ability of learning is high. In other words, this system may be beneficial for very frequent change of products. In terms of assembly quality, 10% to 20% of assembly error (insertion error) is observed in conventional setup (Exp I), while in new cell production setup (Exp II) the error is totally being prevented by the robot assistance, especially by guidance of laser pointer and by the instruction of the assembly sequences.

According to the experimental results, it can be concluded that the developed MASS system

Figure 12. Difference of assembly time by experts and novices



can accelerate the operator's assembly process as well as prevent assembly errors.

According to Zhang (2008), this cable harness task is a kind of cognitive assembly task (Norman, 1993); therefore, the mental work load of the operators cannot be ignored. To evaluate the mental work load of the operators in (Exp I) and (Exp II), NASA-TLX (Task Load Index) method (NASA-TLX for Windows U.S.) was used. After the operators finished the cable harness task, they were required to answer the questionnaires. Based on the NASA-TLX method, the mental work load of the operators can be computed. The mental work load of (Exp I) is 62, which is much higher than that of (Exp II), which is 38. This means that based on the support of MASS, the mental work load of the operators can be reduced significantly.

CONCLUSION

This work aims to realize a new cellular manufacturing system for frequent changes of products. In this chapter, a multi-modal assembly-support system (MASS) was developed for a cellular manufacturing system. In the MASS, two manipulators are used to replace the operators to execute the laborious tasks. Based on the assembly information database and assembly information supporting system, this system is capable of meeting the assembly and training requirements of the experienced and the novice operators. Besides developing the actual system, a simulator for an entire assembly system was created to reduce the time and costs required for development. To protect the operator from harm, several safety strategies and equipments were presented. According to inherent safety theory, two manipulators are ameliorated, which could reduce the injury of the operators even when they were collided by the manipulators.

To evaluate the effect of MASS, a group of experienced operators and novice operators were required to execute a cable harness task. According to the experimental results, basing on the support of MASS, not only the assembly time and the error ratios are reduced, but also the mental work load of the operators is reduced. Therefore, the MASS allows an operator to receive physical and informational support while working in the actual manufacturing assembly process.

Future studies should be directed at identifying and monitoring the conditions that contribute to the operator's fatigue and intention during the assembly process; these efforts will lead to improvements in comfort for the operators and assembly efficiency.

ACKNOWLEDGMENT

This study is supported by NEDO (New Energy and Industrial Technology Development Organization) as one of "Strategic Projects of Element Technology for Advanced Robots". The author Feng DUAN is supported by the Fundamental Research Funds for the Central Universities (No.65010221). We appreciate NEDO and MSTC for accelerating the development of this practical system. In particular, we would like to acknowledge FANUC Company for their excellent cooperation and technical support.

REFERENCES

Arai, T., Duan, F., Kato, R., Tan, J. T. C., Fujita, M., Morioka, M., & Sakakibara, S. (2009). A new cell production assembly system with twin manipulators on mobile base. *Proceeding of 2009 International Symposium on Assembly and Manufacturing* (pp. 149-154). Suwon, Korea.

Bauer, A., Wollherr, D., & Buss, M. (2008). Human-robot collaboration: A survey. *International Journal of Humanoid Robotics*, *5*(1), 47–66. doi:10.1142/S0219843608001303

Beauchamp, Y., & Stobbe, T. J. (1995). A review of experimental studies on human-robot system situations and their design implications. *The International Journal of Human Factors in Manufacturing*, *5*(3), 283–302. doi:10.1002/hfm.4530050306

Colgate, J. E., Wannasuphoprasit, W., & Peshkin, M. A. (1996). Cobots: Robots for collaboration with human operators. *Proceedings of the International Mechanical Engineering Congress and Exhibition*, *58*, 433–439.

Duan, F. (2009). *Assembly skill transfer system for cell production*. Unpublished doctoral dissertation, The University of Tokyo, Japan.

Duan, F., Morioka, M., Tan, J. T. C., & Arai, T. (2008). Multi-modal assembly-support system for cell production. *International Journal of Automation Technology*, *2*(5), 384–389.

Duan, F., Tan, J. T. C., Kato, R., & Arai, T. (2009). Operator monitoring system for cell production. *The International Journal of the Robotics Society of Japan-Advanced Robotics*, *23*, 1373–1391.

Hayakawa, Y., Kitagishi, I., & Sugano, S. (1998). Human intention based physical support robot in assembling work. *Proceedings of the 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 930-935). Victoria, B.C., Canada.

Helms, E., Schraft, R. D., & Hagele, M. (2002). Rob@work: Robot assistant in industrial environments. *Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication* (pp. 399-404). Berlin, Germany.

Ikeda, H., & Saito, T. (2005). *Proposal of inherently safe design method and safe design indexes for human-collaborative robots*. (. *Specific Research Reports of the NIIS-SRR-NO.*, *33*, 5–13.

ISO10218-1. (2006). *Robots for industrial environments- Safety requirements - Part 1: Robot*.

Isa, K., & Tsuru, T. (2002). Cell production and workplace innovation in Japan: Towards a new model for Japanese manufacturing? *Industrial Relations*, *41*(4), 548–578. doi:10.1111/1468-232X.00264

ISO 12100. (2010). *Safety of machinery*. Retrieved from http://www.iso.org/iso/iso_catalogue/ catalogue_ics/catalogue_detail_ics.htm?csnumber=27239

ISO 14121-1. (2007). *Safety of machinery - Risk assessment-Part 1: Principles*.

Kosuge, K., Yoshida, H., Taguchi, D., Fukuda, T., Hariki, K., Kanitani, K., & Sakai, M. (1994). Robot-human collaboration for new robotic applications. *Proceeding of the 20th International Conference on Industrial Electronics, Control, and Instrumentation* (pp. 713-718).

Mayer, R. E. (2001). *Multi-media learning*. New York, NY: Cambridge University Press.

NASA. (n.d.). *NASA-TLX for Windows*. US Naval Research Laboratory. Retrieved from http://www.nrl.navy.mil/aic /ide/NASATLX.php

Neider, J., Davis, T., & Woo, M. (1993). *OpenGL programming guide: The official guide to learning OpenGL*. Addison-Wesley Publishing Company.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Addison Wesley Publish Company.

Oborski, P. (2004). Man-machine interactions in advanced manufacturing systems. *International Journal of Advanced Manufacturing Technology*, *23*(3-4), 227–232. doi:10.1007/s00170-003-1574-5

Reinhart, C., & Patron, C. (2003). Integrating augmented reality in the assembly domain — Fundamentals, benefits and applications. *Annals of the CIRP*, *52*(1), 5–8. doi:10.1016/S0007-8506(07)60517-4

Roboguide, F. A. N. U. C. (n.d.). *Robot system animation tool.* Retrieved from http://www.fanuc.co.jp/en/ product/robot/ro boguide/index.html

Seki, S. (2003). One by one production in the "Digital Yatai" —Practical use of 3D-CAD data in the fabrication. *Journal of the Japan Society of Mechanical Engineering*, *106*(1013), 32–36.

Sugi, M., Nikaido, M., Tamura, Y., Ota, J., Arai, T., & Kotani, K. … Sato, Y. (2005). Motion control of self-moving trays for human supporting production cell "attentive workbench". *Proceedings of the 2005 IEEE International Conference of Robotics and Automation* (pp. 4080-4085). Barcelona, Spain.

Tan, J. T. C., Duan, F., Zhang, Y., Watanabe, K., Pongthanya, N., & Sugi, M. … Arai, T. (2008). Assembly information system for operational support in cell production. *The 41st CIRP Conference on Manufacturing Systems* (pp. 209-212).

Wemmerlov, U., & Johnson, D. J. (1997). Cellular manufacturing at 46 user plants: Implementation experiences and performance improvements. *International Journal of Production Research*, *35*(1), 29–49. doi:10.1080/002075497195966

Zhang, Y., Duan, F., Tan, J. T. C., Watanabe, K., Pongthanya, N., & Sugi, M. … Arai, T. (2008). A study of design factors for information supporting system in cell production. *The 41st CIRP Conference on Manufacturing Systems* (pp. 319-322).

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 412-427, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 33
Modeling and Simulation of Discrete Event Robotic Systems Using Extended Petri Nets

Gen'ichi Yasuda
Nagasaki Institute of Applied Science, Japan

ABSTRACT

This chapter deals with modeling, simulation, and implementation problems encountered in robotic manufacturing control systems. Extended Petri nets are adopted as a prototyping tool for expressing real-time control of robotic systems and a systematic method based on hierarchical Petri nets is described for their direct implementation. A coordination mechanism is introduced to coordinate the event activities of the distributed machine controllers through friability tests of shared global transitions. The proposed prototyping method allows a direct coding of the inter-task cooperation by robots and intelligent machines from the conceptual Petri net specification, so that it increases the traceability and the understanding of the control flow of a parallel application specified by a net model. This approach can be integrated with off-the-shelf real-time executives. Control software using multithreaded programming is demonstrated to show the effectiveness of the proposed method.

1. INTRODUCTION

Complex robotic systems such as flexible manufacturing systems require sophisticated distributed real-time control systems. A major problem concerns the definition of the user tasks and the cooperation between the subsystems, especially since the intelligence is distributed at a low level (machine level). Controlling such systems generally requires a hierarchy of control units corresponding to several abstraction levels. At the bottom of the hierarchy, i.e. machine control level, are the programmable logic controllers (PLC). The next level performs coordination of the PLCs. The third level implements scheduling, that is, the real-time assignments of workpieces

DOI: 10.4018/978-1-4666-1945-6.ch033

and tools to machines. At the machine level, PLCs perform local logical control operations of flexible, modular, high-speed machines through the use of multiple independent drives (Holding, et al. 1992). Implementation languages can be based on ladder diagrams or more recently state machines (Silva, et al. 1982). However, when the local control is of greater complexity, the above kinds of languages may not be well adapted. The development of industrial techniques makes a sequential control system for robotic manufacturing systems larger and more complicated one, in which some subsystems operate concurrently and cooperatively (Neumann, 2007).

In the area of real-time control of discrete event robotic systems the main problems that the system designer has to deal with are concurrency, synchronization, and resource sharing problems. Presently, the implementation of such control systems makes a large use of microcomputers. Real-time executives are available with complete sets of synchronization and communication primitives. However, coding the specifications is a hazardous work and debugging the implementation is particularly difficult when the concurrency is important. It is important to have a formal tool powerful enough and allowing developing validation procedures before implementation. Conventional specification languages do not allow an analytical validation. Consequently, the only way to validate is via simulation and step by step debugging. On the other hand, a specification method based on a mathematical tool may be more restricting, but analytical procedures can strongly reduce the simulation step. Rapid prototyping is an economical and crucial way to experiment, debug, and improve specifications of parallel applications. Increasing complexity of synchronizing mechanisms involved in concurrent system design makes necessary a prototyping step starting from a formal, already verified model. Petri nets allow to validate and to evaluate a model before its implementation. The formalism allowing a validation of the main properties of the Petri net control structure (liveness, boundedness, etc.) guarantees that the control system will not fall immediately in a deadlocked situation. In the field of flexible manufacturing cells, the last aspect is essential because the sequences of control are complex and change very often.

When using Petri nets, events are associated with transitions. Activities are associated to the firing of transitions and to the marking of places which represents the state of the system. The network model can describe the execution order of sequential and parallel tasks directly without ambiguity (Silva, 1990). Pure synchronization between tasks, choices between alternatives and rendezvous can be naturally represented. Moreover at the machine level Petri nets have been successfully used to represent the sequencing of elementary operations (Yasuda, et al. 1992). In addition to its graphic representation differentiating events and states, Petri nets allow the possibility of progressive modeling by using stepwise refinements or modular composition. Libraries of well-tested subnets allow components reusability leading to significant reductions in the modeling effort. The possibility of progressive modeling is absolutely necessary for large and complex systems, because the refinement mechanism allows the building of hierarchically structured net models. Furthermore, a real-time implementation of the Petri net specification by software called a token player can avoid implementation errors. This is because the specification is directly executed by the token player and the implementation of these control sequences preserves the properties of the model. In this approach, the Petri net model is stored in a database and the token player updates the state of the database according to the operation rules of the model. For control purposes, this solution is very well suited to the need of flexibility, because, when the control sequences change, only the database needs to be changed.

Some techniques derived from Petri nets have been successfully introduced as an effective tool for describing control specifications and realizing the control in a uniform manner (Murata, et al. 1986). However, in the field of flexible manufacturing cells, the network model becomes complicated and it lacks the readability and comprehensibility (David, et al. 1992). Therefore, the flexibility and expandability are not satisfactory in order to deal with the specification change of the control system. Despite the advantages offered by Petri nets, the synthesis, correction, updating, etc. of the system model and programming of the controllers are not simple tasks (Zhou, et al. 1993), (Desrochers, et al. 1995), (Lee, et al. 2006). Some Petri net implementation methods have already been proposed for simulation purposes or for application prototyping (Butler, 1991), (Garcia, 1998), (Piedrafita, et al. 2008). However the implementation method for hierarchical and distributed control of complex robotic systems has not been established sufficiently so far (Breant, et al. 1992), (Girault, et al. 2003), (Zhou, et al. 1999). If it can be implemented using Petri nets, the modeling, simulation and control can be realized consistently.

This chapter describes a Petri net based prototyping method for real-time control of complex robotic systems. The presented method, based on the author's previous works (Yasuda, et al. 2010), (Yasuda, 2010), involves three major steps, and progressively gathers all information needed by the control system design and the code generation for simulation experiments. The first step consists in specifying the conceptual net model for overall system control. The second step consists in transforming the net model into the detailed net model. Based on the hierarchical and distributed structure of the system, the specification procedure is a top-down approach from the conceptual level to the detailed level. The third step consists in decomposing the detailed net into local net models for machine control and the coordination model. The coordination algorithms are simplified since the robots and machines in the system are separately controlled using dedicated task execution programs. In order to deal with complex models, a hierarchical approach is adopted for the coordination model design. In this way, the macro representation of the system is broken down to generate the detailed nets at the local machine control level. Finally, the C++ code generation using multithreaded programming is described for the prototype hierarchical and distributed control system.

2. MODELING OF DISCRETE EVENT SYSTEMS USING EXTENDED PETRI NETS

A Petri net is a directed graph whose nodes are places shown by circles and transitions shown by bars. Directed arcs connect places to transitions and transitions to places. Formally, a Petri net is a bipartite graph represented by the 4-tuple $G=\{P,T,I,O\}$ such that:

$P=\{p_1,p_2,\ldots,p_n\}$ is a finite, not empty, set of places;
$T=\{t_1,t_2,\ldots,t_m\}$ is a finite, not empty, set of transitions;
$P\cap T=\phi$, i.e. the sets P and T are disjointed;
I: $T\rightarrow P^\infty$ is an input function, a mapping from transitions to bags of places;
O: $T\rightarrow P^\infty$ is an output function, a mapping from transitions to bags of places.

The input function I maps from a transition t_j to a collection of places $I(t_j)$, known as input places of a transition. The output function O maps from a transition t_j to a collection of places $O(t_j)$, known as input places of a transition. The pre-incidence matrix of a Petri net is $C^- = [c_{ij}^-]$ where $c_{ij}^- = 1$ $(p_i \in I(t_j))$, $c_{ij}^- = 0$ $(p_i \notin I(t_j))$; the post–incidence matrix is $C^+ = [c_{ij}^+]$ where $c_{ij}^+ = 1$ $(p_i \in O(t_j))$, $c_{ij}^+ = 0$ $(p_i \notin O(t_j))$, then the incidence matrix of the Petri net $C = C^+ - C^-$. Each place contains integer (positive or zero) marks or tokens. The number of tokens in each place is defined by the marked vector or marking $M = (m_1, m_2, ..., m_n)^T$. The number of tokens in one place p_i is simply indicated by $M(p_i)$. The firing of a transition will change the token distribution (marking) in the net according to the transition firing rule. In the basic Petri net, a transition t_j is enabled if, $\forall p_i \in I(t_j)$, $M_k(p_i) \geq w(p_i, t_j)$, where the current marking is M_k and $w(p_i,t_j)$ is the weight of the arc from p_i to t_j. A sequence of firings will result in a sequence of markings. A marking M_n is said to be reachable from a marking M_0 if there exists a sequence of firings that transforms M_0 to M_n. The set of all possible markings reachable from M_0 is denoted as $R(M_0)$. A Petri net is said to be k-bounded or simply bounded if the number of tokens in each place does not exceed a finite number for any marking reachable from M_0, i.e., $\forall p_i \in P$, $\forall M \in R(M_0)$, $M(p_i) \leq k$ (Reisig, 1985), (Murata, 1989).

In the basic Petri net, bumping occurs when despite the holding of a condition, the preceding event occurs. This can result in the multiple holding of that condition. From the viewpoint of discrete event process control, bumping phenomena should be excluded. So, the firing rule was modified so that the system is free of this phenomenon. Because the modified Petri net must be 1-bounded, for each place p_i, $m_i = 0$ or 1, and the weight of every arc is 1. A Petri net is said to be ordinary if all of its arc weights are 1's. Thus the axioms of the modified Petri net are as follows.

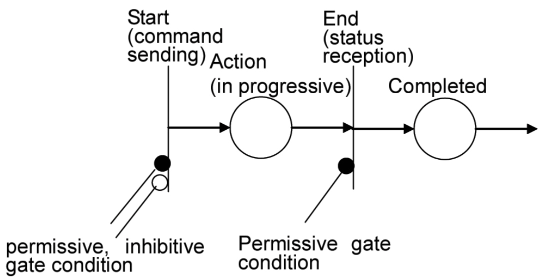
1. A transition t_j is enabled if for each place $p_k \in I(t_j)$, $m_k = 1$ and for each place $p_l \in O(t_j)$, $m_l = 0$.
2. When an enabled transition t_j fires, the marking M is changed to M', where for each place $p_k \in I(t_j)$, $m'_k = 0$ and for each place $p_l \in O(t_j)$, $m'_l = 1$.
3. In any initial marking, there must not exist more than one token in each place.

A transition without any input place is called a source transition, and one without any output place is called a sink transition. A source transition is unconditionally enabled, and the firing of a sink transition consumes a token in each input place, but does not produce any. According to these axioms, the number of tokens in a place never exceeds one, thus the modified Petri net is essentially 1-bounded and said to be a safe graph.

Besides the guarantee of safeness, considering not only the modeling but also the actual system control of robotic systems, the additional capabilities of input and output interfaces which connect the net to its environment are required. The extended Petri net adopts the following two elements: 1) Gate arc, 2) Output signal arc (Hasegawa, et al. 1984). A gate arc connects a transition with a signal source, and depending on the signal, it either permits or inhibits the occurrence of the event which corresponds to the connected transition. Gate arcs are classified as permissive or inhibitive, and internal or external. An output signal arc sends a command request signal from a place to an external machine. The interfaces are a set of transitions which represent the communication activities of the net with its environment. Thus the firing rule of a transition is described as follows: an enabled transition may fire when it does not have any internal permissive arc signaling 0 nor any internal inhibitive arc signaling 1 and it does not have any external permissive arc signaling 0 nor any external inhibitive arc signaling 1.

A robotic action is modeled by two transitions and one condition as shown in Figure 1. At the "Start" transition the command associated with the transition is sent to the corresponding robot or machine. At the "End" transition the status report is received. When a token is present in the "Action" place, the action is in progressive. The "Completed" place can be omitted, and then the "End" transition is fused with the "Start" transi-

Figure 1. Extended Petri net model of robotic action with external permissive and inhibitive gate arcs



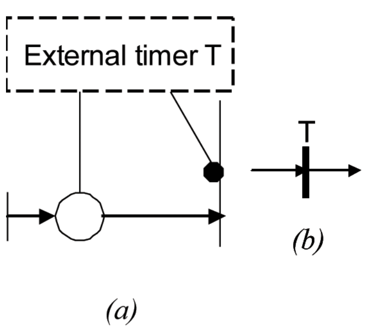
tion of the next action. Activities can be assigned an amount of time units to monitor them in time for real performance evaluation.

The firing of a transition is indivisible; the firing of a transition has duration of zero. Extended Petri nets to consider timing conditions where each activity is assigned an amount of time units can also be used as shown in Figure 2. Through the simulation steps, the transition vector table is efficiently used to extract enabled or fired transitions.

The flow chart of simulation and evaluation procedure is shown in Figure 3. At a step of simulation of robotic task, the configuration of the robots can be seen with graphic simulation. The data structure of the extended Petri net simulator is made up of several tables corresponding to the structural information of the net specifying the robotic task. These tables are the following ones:

1. The table of the labels of the input and output places for each transition,
2. The table of the transitions which are likely to be arbitrated for each conflict place,
3. The table of the gate arcs which are internal or external, permissive or inhibitive, for each transition,
4. The table of marking which indicates the current marking for each place,
5. The table of places to tasks mapping which points out the tasks that have to be put into the ready state when the corresponding place receives a token,
6. The table of the "end of task" transitions, which associates with each of task the set of transitions with external gate arcs switched each time an "end of task" message is received by the simulator. The "end of task" transitions are only fired on the reception of an "end of task" message.

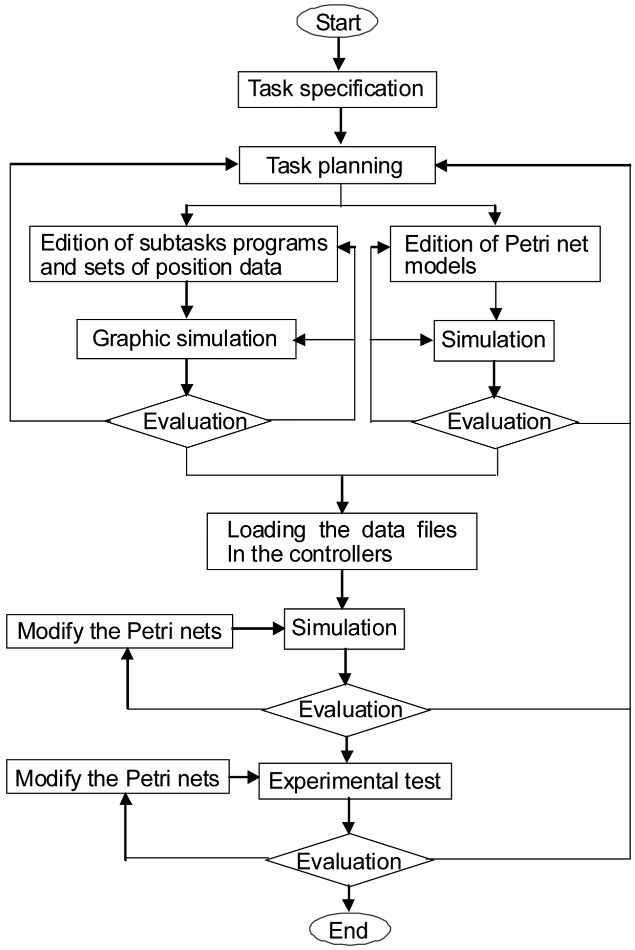
Figure 2. Examples of representation of timing conditions: external timer with output signal arc and external gate arc, (b) timed transition



3. PETRI NET MODELS OF COOPERATIVE CONTROL OF CONCURRENT TASKS

A robotic task consists of several subtasks or operations, and a subtask consists of several actions. Conceptually, robotic processes are represented as sequential constructs or state machines, where each transition has exactly one incoming arc and exactly one outgoing arc. The structure of a place having two or more output transitions is referred to as a conflict, decision, or choice, depending on applications. State machines allow the representation of decisions, but not the synchronization of parallel activities. In a net model of robotic task, the set of places can be classified into three groups: idle, operation, and resource places. A token in an idle place indicates that the robot is ready to work and

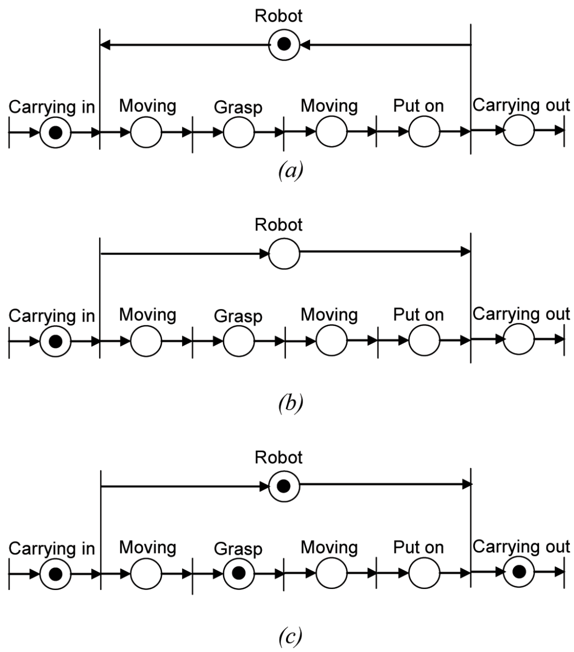
Figure 3. Flow chart of simulation and evaluation procedure



waiting for a specified signal from another robot or its environment. An operation place represents an operation to be processed for a workpiece or part in a manufacturing sequence and initially it has no token. Resource places represent resources (robots and machines) and their initial tokens represent the number of available resource units.

A task executed by a robot or machine can be seen as some connection of more detailed subtasks. For example, transferring an object from a start position to a goal position is a sequence of the following subtasks: moving the hand to the start position, grasping the object, moving to the goal position, and putting it on the specified place. Figure 4 shows the net representation of a robotic task: pick and place operation with the input and output conveyors. While the place "Robot" in Figure 4(a) indicates that the state of the robot is "ready" when the token is in the place, in Figure 4(b) it indicates that the state of the robot is "operating". The place also indicates the macro representation of the pick and place operation. The parallel net in Figure 4(b) is equivalent to the cyclic net in Figure 4(a) with respect to the enabling conditions of all the transitions in the net. The parallel net assures that the robot can load or unload only one workpiece at a time. Figure 4(c) shows the possible evolution of dynamic behavior of the net.

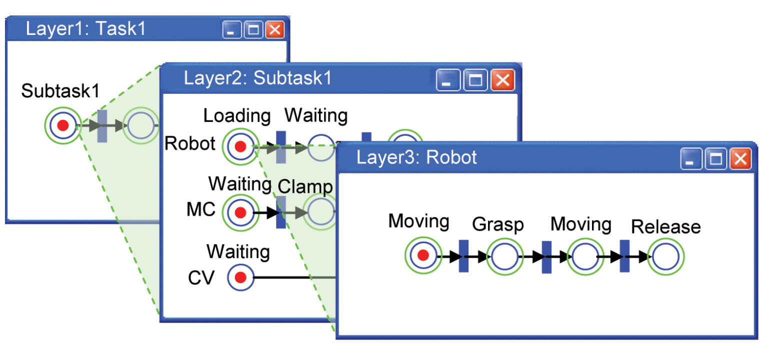
Figure 4. Net representation of robotic task: (a) cyclic net model of robot, (b) equivalent parallel net model, (c) possible state of net model



Furthermore the subtasks are translated into more detailed actions. A hierarchical approach consists in building a model by stepwise refinements. At each step some parts of the model are replaced by a more complex model. A modular approach by model composition affords to build complex validated Petri nets. Figure 5 shows the view of a hierarchical net representation by the graphic net simulator.

A specification procedure for discrete event robotic systems based on Petri nets is as follows. First, the conceptual level activities of the system are defined through a net model considering the task specification corresponding to the aggregate discrete event process. The places which represent the subtasks indicated as the task specification are connected by arcs via transitions in the specified order corresponding to the flow of subtasks and a workpiece. The places representing

Figure 5. Hierarchical net representation of robotic task (loading) on graphic net simulator



robots and machines used for the subtasks are also added to connect transitions which correspond to the beginning and ending of their subtasks. Then, the places describing the subtasks are substituted by a subnet based on activity specification and required control strategies in a manner which maintains the structural properties (Yasuda, et al. 2010).

For concurrent control in the conceptual net model, two implementation methods of synchronous interaction between two tasks which are executed respectively by one robot are shown in Figure 6(a), (b) (Yasuda, 2000), while an implementation with asynchronous communication based on the well-known signal/wait (semaphore) concept is shown in Figure 6(c).

A coordination mechanism is introduced to coordinate concurrent systems with separate robots which require interaction between each other, such as synchronization and resource conflict. Figure 7 shows the net model of a coordination mechanism to conduct synchronous interaction by means of synchronous communication between two robots. This is also the detailed representation of the shared transition. A shared transition for synchronous interaction by separate robots is said to be a global transition, while a transition for independent action by a single robot is said to be a local transition. For synchronous interaction, the coordination algorithm is formally expressed using logical variables such that the global transition is fired if all of associated transitions in the local net models are fired. The firing condition of a global transition G_j in the conceptual net which represents the event of synchronous action by S robots is written as

$$G_j = \bigcap_{sub=1}^{S} t_{jsub} \tag{1}$$

where the corresponding event of action by each robot is represented by t_{jsub} (sub = 1⋯S), and ∩ denotes logical product operation. Figure 8 shows a net model of a coordination mechanism to execute selective control by means of synchronous communication, where the decision place executes any arbitration rule such as order or priority, to select independent action by a robot or cooperative action by two robots.

A hierarchical and distributed control system is composed of one system controller and several machine controllers. The coordination mechanism as well as the conceptual net model of the system is implemented in the system controller and detailed net models are allocated to machine controllers. The coordination program is substantially the firing test program of global transitions using the firing rule and a set of relational data tables of global transitions with their input and output places and associated gate arcs. The coordination procedure through communica-

Figure 6. Net representation of synchronous interaction between two concurrent tasks: (a) synchronous communication with a shared transition, (b) interlock with mutual gate arcs,(c) asynchronous communication (signal/wait)

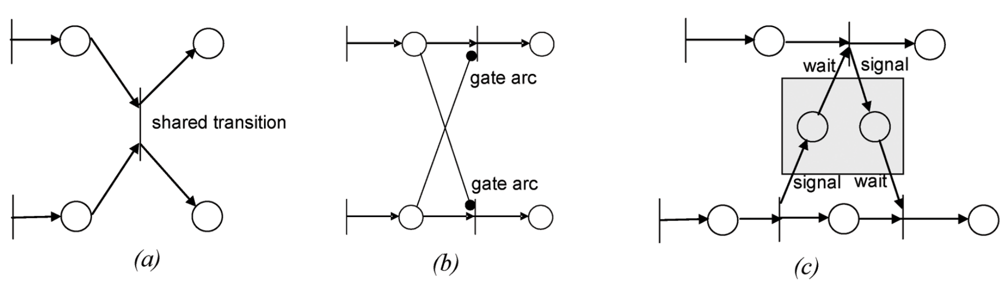
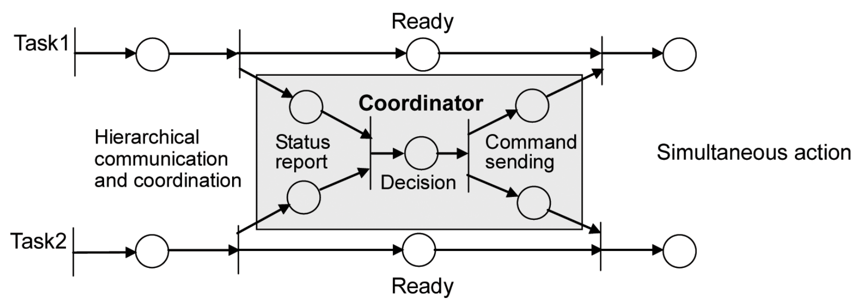


Figure 7. Net model of coordination mechanism for synchronous interaction by means of synchronous communication with a shared transition



tion between the coordinator and machine controllers is shown as follows:

1. When a machine controller receives a command start signal from the coordinator, it starts the execution of the requested command task.
2. At the end of execution, the machine controller sends a status report as a gate condition to the coordinator.
3. When the coordinator receives the new gate condition, it updates the net model and tests the firing condition associated with the gate condition in the net. If a new transition is fired, then a command associated with the transition is sent to the corresponding machine controller.

For the actual control, the operations of each machine or robot are broken down into a series of unit motions, which is represented by mutual connection between places and transitions. A place means a concrete unit motion of a machine. From these places, output signal arcs are connected to the external machines, and external gate arcs from the machines are connected to the transitions of the net when needed, for example, to synchronize and coordinate operations. When a token enters a place that represents a subtask, the machine defined by the machine code is informed to execute a specified subtask with positional and control data; these code and data are defined as the place parameters. Figure 9 shows the net representation of real-time execution control of a robotic unit action.

Figure 8. Net model of coordination mechanism for selective control

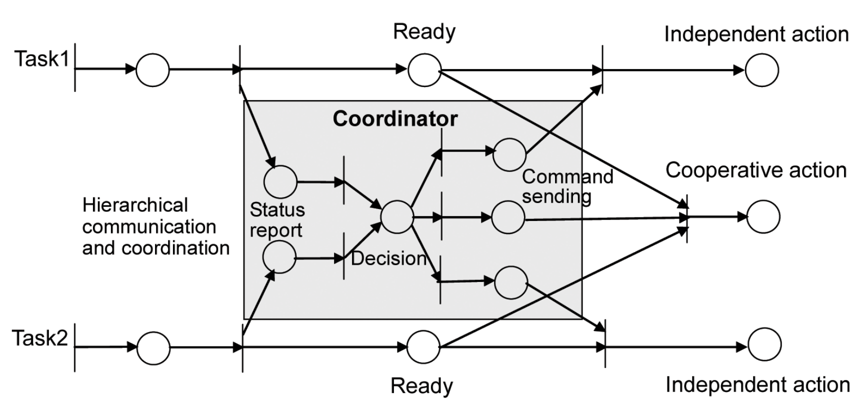
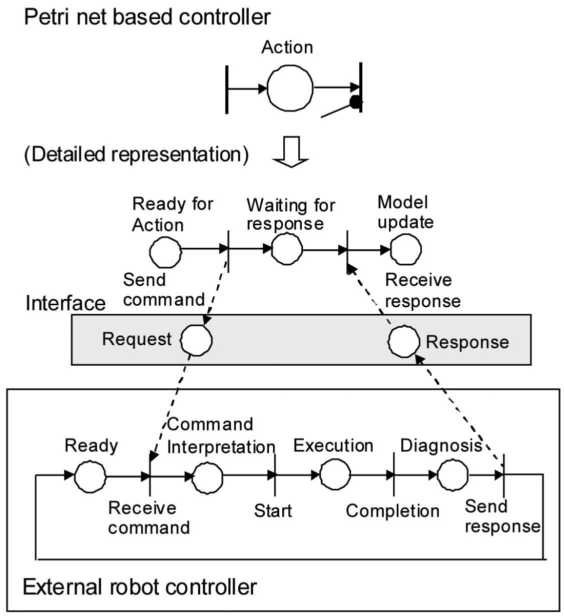


Figure 9. Net representation of execution control of robotic action using output signal arc and external permissive gate arc



4. IMPLEMENTATION OF DISTRIBUTED CONTROL WITH MULTITHREADS

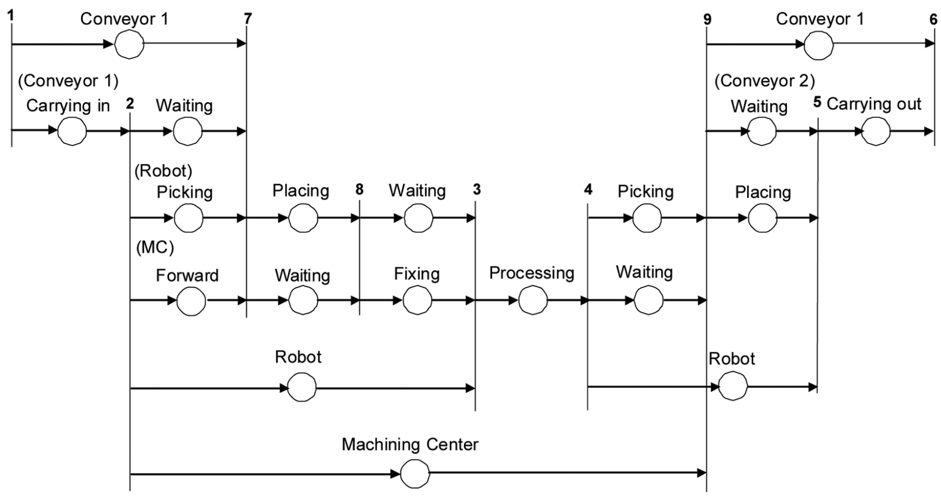
The example robotic system has one robot, one machining center, and two conveyors, where one is for carrying in and the other one is for carrying out. The main execution of the system is indicated as the following task specification:

1. A workpiece is carried in by the input conveyor.
2. The workpiece is loaded to the machining center by the robot.
3. The workpiece is processed by the machining center.
4. The workpiece is unloaded from the machining center by the robot.
5. The workpiece is carried out by the output conveyor.

The robotic system works in the following way: A workpiece comes on the input conveyor up to the take up position. The robot waits in the waiting position in front of the conveyor, and, on stopping, approaches in the take up position, grips the object and returns to the waiting position. Then it turns, goes into the working space of the machine center and there it leaves the workpiece. After automatic gripping of the object the robot draws back and it waits for the machining center to complete object processing. After object processing, the robot goes to the machining center, takes the workpiece from an opened vice and carries it over to a free position on the output conveyor.

The discrete event processes of the robotic system are represented at the conceptual level as shown in Figure 10. All the transitions are used to coordinate the overall system. Shared transitions between the robot and a machine, represent synchronous interaction and are coordinated as global transitions to be fired simultaneously. In this step, if necessary, control conditions such as the capacity of the system between the respective subtasks must be connected to regulate the execution of the manufacturing process. Next, each place

Figure 10. Net representation of robotic task at conceptual level



representing a subtask at the conceptual level is translated into a detailed subnet, which is used for local control of each robot or machine.

The prototyping method was applied to produce semi-automatically C++ program from the net models on a general PC using multithreaded programming (Grehan, et al. 1998). Then, by executing the coordination program and net based controllers algorithms based on loaded information on a network of dedicated microcomputers, experiments on the final test can be performed (Yasuda, 2010). Multithreads control software composed of one modeling thread, one simulation thread and several task execution threads, were designed and written in Microsoft Visual C# under OS Windows XP SP3. The simulation thread executes the coordination program and the conceptual net based controller algorithm, while the task execution threads execute local net based controller algorithms, which control robots and machines through serial interfaces using the command/response concept. An example diagram of two-level net based concurrent real-time control of two external machines using one simulation and two task execution threads is shown in Figure 11.

The modeling thread, which is the main thread, executes the event driven net modeling, drawing and modification based on task specification using windows button clicks and mouse operations, as shown in Figure 12. When the transformation of graphic data of the net model into internal structural data is finished, the simulation thread is activated using window buttons by the user from the modeling thread. The simulation thread executes the enabling and firing test using gate conditions as shown in Figure 13, and when a transition is fired, the simulation thread activates the task execution thread and initiates the execution of a subtask by sending commands attached to the fired transition. When all the subtasks in the system are in progressive, the simulation thread waits for the turning on of any gate condition repeatedly. If a subtask is completed, the gate condition is turned on, then the simulation thread receives the gate signal through the external gate arc and updates the table of gate conditions.

The program structure and the main C# code of the task execution thread are illustrated in Figure 14 and Listing 1. One task execution thread is allocated in each machine. When a task execution thread is activated, it sends the direct commands with specified positional and control data

Figure 11. Two-level net based concurrent real-time control of two external machines

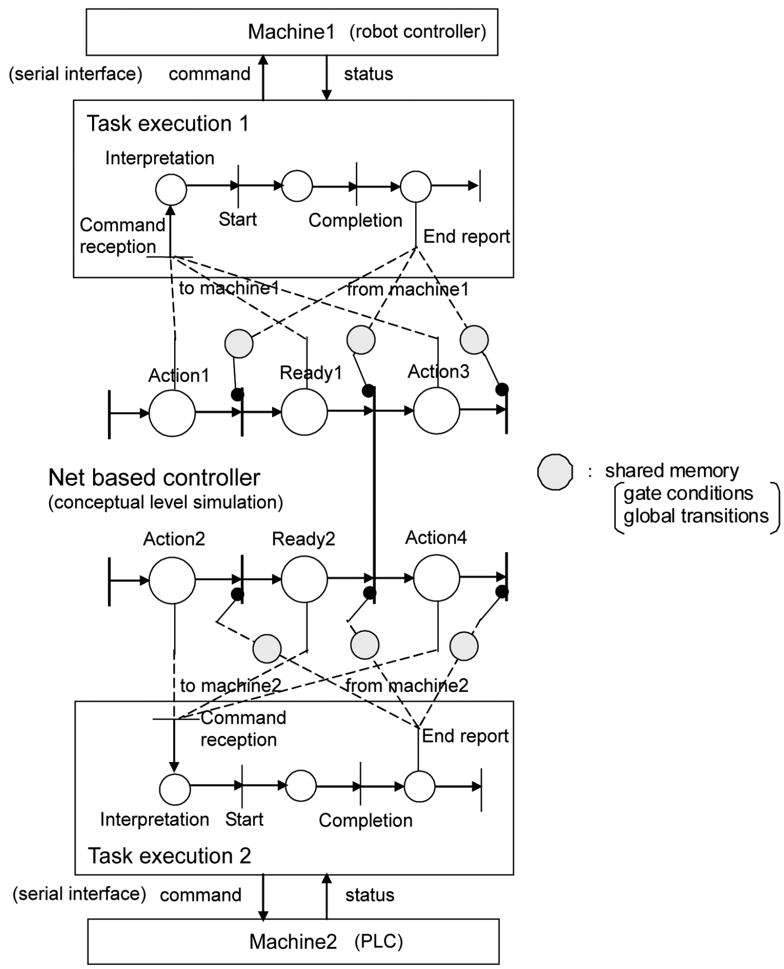


Figure 12. Program structure and main C# code of modeling thread

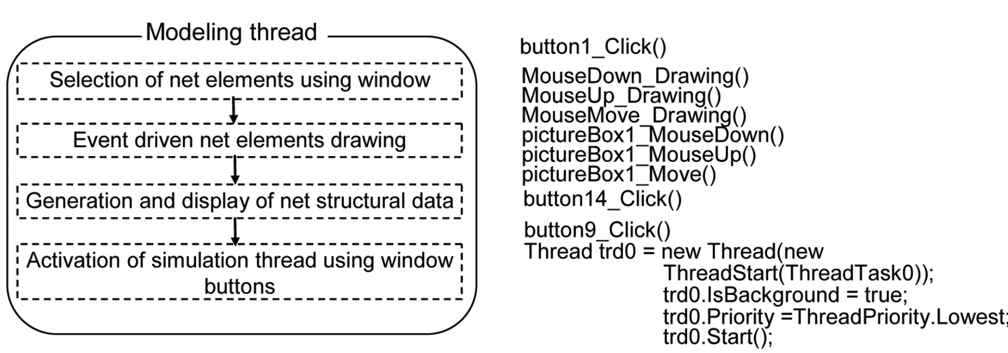
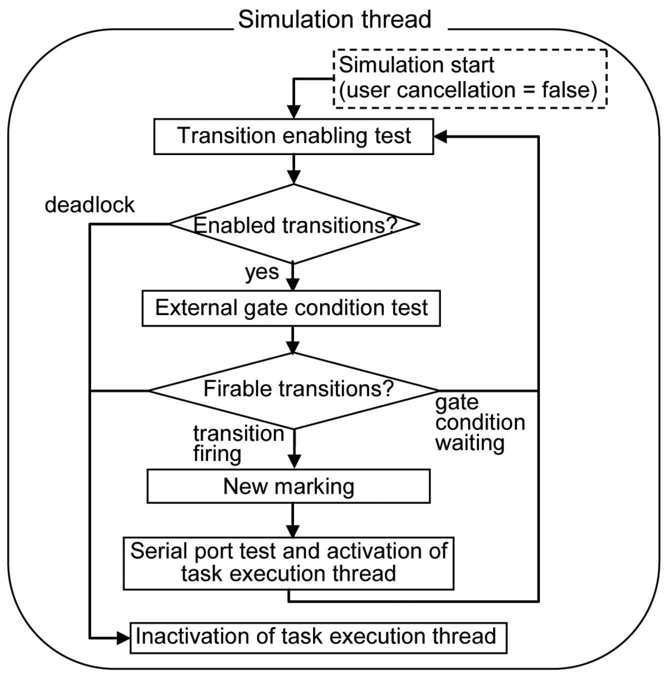


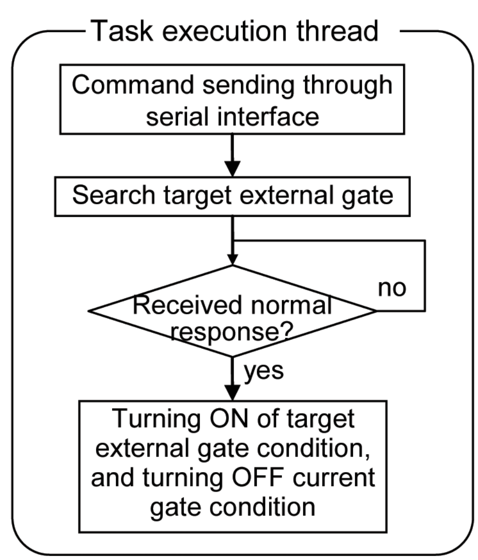
Figure 13. Program structure of simulation thread



through the serial interface to the dedicated robot or machine. Then the thread searches the target gate condition and waits for any status report through the interface repeatedly. When the subtask ends its activity normally, the thread receives the normal end report. Then, the thread turns the target gate condition on and the current gate condition off, so that the simulation thread can proceed with the next activations through the external gate arc. During simulation and task execution, it is decided by the simulator that whether it is in a deadlocked situation or not, and the user can stop the task execution through the simulator at any time.

In the real-time control both the simulation thread and the task execution threads access to the external gate variables as shared variables; task execution threads write the values of the gate variables after the completion of subtasks and the simulation thread reads them for the friability test considering gate conditions. Mutual exclusive access control and the C# code were implemented as shown in Figure 15 and Listing 2 respectively. The method function uses a "lock" statement of C# for the mutual exclusive access to shared

Figure 14. Program structure of task execution thread



Listing 1. Task execution thread

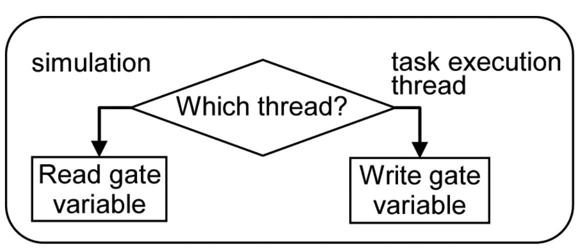
```
private void ThreadTask1()
{
        serialPort1.WriteLine(CC1[TH1 - 1] + "\r\n");        // command of fired transition no.TH1
        for (int j = 0; j < TP_R; j++){
                if (TPC[TH1 - 1, j] > 0){                    // search output place of fired transition
                        PO = TPC[TH1 - 1, j];
                        PH[PO - 1] = 1;  }}                  // output place no.PO
                        for (int i = 0; i < T_No; i++){      // search output transition of place no.PO
                                for (int j = 0; j < TP_R; j++){
                                        if (TPC[i, j] == -(PO + 1)){
                                                for (int m = 0; m < EG_R; m++){
                                                        if (TGE[i, m] > 0)  GC1 = TGE[i, m];     // gate no.GC1
                                                }}}
                                        do{
                                                // method call
                                        }
                                        while (THN == 1);                // gate condition write mode on
                        }
        if (THN == 0) receiveData1 = "EG-WAIT";
}
```

variables so that, while one thread calls the function, the other thread can not call it. Using the method function call, the simulator thread waits for the external gate signal, and after a task execution thread writes the permissive value of the target gate variable, the simulator thread reads it.

Experimental results of multithreads scheduling for one and two task execution threads are shown in Figure 16(a), (b), respectively. In case of two threads, one thread takes charge of a robot controller, and another takes charge of a PLC for sequence control of a conveyor respectively through serial interface. Here, the time slice of the OS is about 15ms, and the timer program is inserted in the reference position of each thread to capture the time the method function is called.

Numerous experiments show that the gate condition is transferred through the shared memory from the task execution thread to the simulation thread as quickly as possible. Experiments using a real robot and conveyors show that the simulation thread and the task execution threads are concurrently executed with even priority, and the values of external gate variables are changed successfully in the conceptual net model. The global transitions fire simultaneously with the transition of the conceptual net of the whole system task. The robot cooperated with the conveyors and the machining center, and the example robotic system performed the task specification successfully.

Figure 15. Program structure of method function for mutual exclusive access

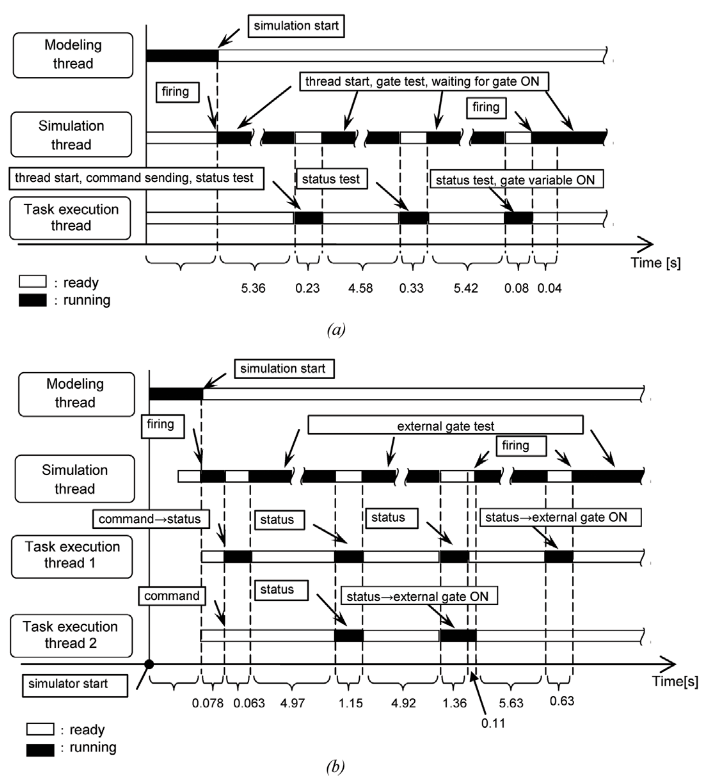


Listing 2. Method function for mutual exclusive access

```
private void EG_SET()
{
        lock (this)
        {
        if (EG_f == 1){                              // simulation thread
                for (int i = 0; i < T_No; i++){      // gate condition test
                if (TI[i] == 1){                     // enabled transition
                        for (int j = 0; j < EG_R; j++){
                                GC1 = TGE[i, j];     // gate no.j associated with transition no.i
                                if (GC1 > 0){        // permissive gate
                                        if (EG[(GC1 - 1)] == 0) TI[i] = 0;       // read permissive gate signal off
                                }
                                else if (GC1 < 0) {  // inhibitive gate
                                        if (EG[Math.Abs(GC1 + 1)] == 1) TI[i] = 0;   // read inhibitive gate signal on
                                        }}}}}

        else if (EG_f == 2){                         // task execution thread
                if (recieveData1 == "RDY"){          // received status report 'ready'
                        GC2 = TGE[TH1-1, j];         // write gate no.GC2 off
                        if (GC2 > 0) EG[GC2 - 1] = 0;
                        else if(GC2 < 0) EG[Math.Abs(GC2 + 1)] = 0;
                        EG[GC1 - 1] = 1;             // write gate no.GC1 on
                        TH[TH1 - 1] = 0;             // fired teansition no. initialize
                        PH[k] = 0;                   // working place no. initialize
                        THN = 0;                     // gate condition write mode off
                }
                Application.DoEvents();              // serial communication message buffer initialize
        }}
```

Figure 16. Experimental results of multithreads scheduling: one task execution thread, (b) two task execution threads



In accordance with the implementation using multithreaded programming, hierarchical and distributed implementation under a real-time operating system on a network of microcomputers connected via a shared serial bus is also possible, where each microcomputer is dedicated to the local net model of a subsystem in the overall robotic system. After the arrival of a request, the response carrying the status information crosses the communication network and gets to the input buffer of the network board of the controller and written in the cache memory. The control data issued by the controller is written in the cache memory shared with the network board, sent to its target microcomputer, and there causes the new controlled status. By a shared communication network replacing traditional directly wired systems, a more flexible, reliable and efficient control performance can be surely expected.

5. CONCLUSION

A prototyping methodology to build hierarchical and distributed control systems corresponding to the hardware structure of robotic control systems has been presented. The conceptual net is used to coordinate distributed local machine controllers using the decomposed information of global transitions representing cooperative interaction between machines: the coordination mechanism can be implemented in each layer of the control hierarchy of the system repeatedly.

The overall control structure of the example robotic system was implemented on a general PC with serial interface using multithreaded programming. For the example system, detailed net models can be automatically generated using the database of robotic operations. The hierarchical approach allows us to reuse validated net models, such as loading, unloading, and specific handling operations, already defined for other purposes, that is an efficient way to deal with complex net models. The conceptual and local net models are not so large that all of the net based controllers can be implemented on general microcomputers or PLCs. Thus, modeling, simulation and control of large and complex manufacturing systems can be performed consistently using Petri nets.

REFERENCES

Breant, F., & Paviot-Adet, E. (1992). OCCAM prototyping from hierarchical Petri nets. In Becker, M., Litzler, L., & Trehel, M. (Eds.), *Transputers '92 Advanced Research and Industrial Applications* (pp. 189–209). Amsterdam, Netherlands: IOS Press.

Butler, B., Esser, R., & Mattmann, R. (1991). A distributed simulator for high order Petri nets. In Rozenberg, G. (Ed.), *Advances in Petri Nets LNCS 483* (pp. 47–63). Berlin, Germany: Springer-Verlag.

Caloini, A., Magnani, G., & Pezze, M. (1998). A technique for designing robotic control systems based on Petri nets. *IEEE Transactions on Control Systems Technology*, *6*(1), 72–87. doi:10.1109/87.654878

David, R., & Alla, H. (1992). *Petri Nets and Grafcet: Tools for modelling discrete events systems*. UK: Prentice-Hall International.

Desrochers, A. D., & Al-Jaar, R. Y. (1995). *Applications of Petri Nets in manufacturing systems: Modeling, control and performance analysis*. New York, NY: IEEE Press.

Garcia, F. J., & Villarroel, J. L. (1998). Decentralized implementation of real-time systems using time Petri nets. Application to mobile robot control. In *Proceedings of the 5th IFAC Workshop on Algorithms and Architectures for Real-time Control*, (pp. 11-16).

Girault, C., & Valk, R. (2003). *Petri Nets for systems engineering*. Berlin, Germany: Springer-Verlag.

Grehan, R., Moote, R., & Cyliax, I. (1998). *Real-time programming: A guide to 32-bit embedded development*. Reading, MA: Addison-Wesley.

Hasegawa, K., Takahashi, K., Masuda, R., & Ohno, H. (1984). Proposal of mark flow graph for discrete system control. *Transactions of SICE*, *20*(2), 122–129.

Holding, D. J., & Sagoo, J. S. (1992). A formal approach to the software control of high-speed machinery. In G. W. Irwin & P. J. Fleming (Eds.), *Transputers in real-time control,* (pp. 239-282). Taunton, Somerset, UK: Research Studies Press.

Lee, E. J., Togueni, A., & Dangoumau, N. (2006). A Petri net based decentralized synthesis approach for the control of flexible manufacturing systems. In *Proceedings of the 2006 IMACS Multiconference Computational Engineering in Systems Applications*.

Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, *77*(4), 541–580. doi:10.1109/5.24143

Murata, T., Komoda, N., Matsumoto, K., & Haruna, K. (1986). A Petri net based controller for flexible and maintainable sequence control and its application in factory automation. *IEEE Transactions on Industrial Electronics*, *33*(1), 1–8. doi:10.1109/TIE.1986.351700

Neumann, P. (2007). Communication in industrial automation – What is going on? *Control Engineering Practice*, *15*(11), 1332–1347. doi:10.1016/j.conengprac.2006.10.004

Piedrafita, R., Tardioli, D., & Villarroel, J. L. (2008). Distributed implementation of discrete event control systems based on Petri nets. In *Proceedings of the 2008 IEEE International Symposium on Industrial Electronics,* (pp. 1738-1745).

Reisig, W. (1985). *Petri Nets*. Berlin, Germany: Springer-Verlag.

Silva, M. (1990). Petri nets and flexible manufacturing. In Rozenberg, G. (Ed.), *Advances in Petri Nets LNCS 424* (pp. 374–417). Berlin, Germany: Springer-Verlag.

Silva, M., & Velilla, S. (1982). Programmable logic controllers and Petri nets*:* A comparative study. In Ferrate, G., & Puente, E. A. (Eds.), *In IFAC Software for Computer Control 1982* (pp. 83–88). Oxford, UK: Pergamon.

Yasuda, G. (2000). A distributed control and communication structure for multiple cooperating robot agents. In *IFAC Artificial Intelligence in Real Time Control 2000* (pp. 283–288). Oxford, UK: Pergamon.

Yasuda, G. (2010). Distributed cooperative control of industrial robotic systems using Petri net based multitask processing. In H. Liu, H. Ding, Z. Xiong, & X. Zhu (Eds.), *Proceedings of the 3rd International Conference on Intelligent Robotics and Applications LNAI 6425*, (pp. 32-43). Berlin, Germany: Springer-Verlag.

Yasuda, G., & Ge, B. (2010). Petri net model based specification and distributed control of robotic manufacturing systems. In *Proceedings of the 2010 IEEE International Conference on Information and Automation*, (pp. 699-705).

Yasuda, G., & Tachibana, K. (1992). Implementation of real-time control schemes on a parallel processing architecture using transputers. In *Proceedings of IEEE Singapore International Conference on Intelligent Control and Instrumentation*, (pp. 760-765).

Zhou, M., DiCesare, F., & Desrochers, A. A. (1993). *Petri Net synthesis for discrete event control of manufacturing systems*. London, UK: Kluwer.

Zhou, M., & Venkatesh, K. (1999). *Modeling, simulation, and control of flexible manufacturing systems: A Petri Net approach*. Singapore: World Scientific.

This work was previously published in Prototyping of Robotic Systems: Applications of Design and Implementation, edited by Tarek Sobh and Xingguo Xiong, pp. 51-69, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 34
Human-Friendly Robots for Entertainment and Education

Jorge Solis
Waseda University, Japan & Karlstad University, Sweden

Atsuo Takanishi
Waseda University, Japan

ABSTRACT

Even though the market size is still small at this moment, applications of robots are gradually spreading out from the manufacturing industrial environment to face other important challenges, like the support of an aging society and to educate the new generations. The development of human-friendly robots drives research that aims at autonomous or semi-autonomous robots that are natural and intuitive for the average consumer to interact with, communicate with, and work with as partners, besides learning new capabilities. In this chapter, an overview of research done on the mechanism design and intelligent control strategies implementation on different platforms and their application to entertainment and education domains will be stressed. In particular, the development of an anthropomorphic saxophonist robot (designed to mechanically reproduce the organs involved during saxophone playing) and the development of a two-wheeled inverted pendulum (designed to introduce the principles of mechanics, electronics, control, and programming at different education levels) will be presented.

INTRODUCTION

The development of anthropomorphic robots is inspired by the ancient dream of humans replicating themselves. However, human behaviors are difficult to explain and model. The recent technological advances in robot technology, artificial intelligence, power computation, etc. have contributed to enable humanoid robots to roughly emulate the physical dynamics and motor dexterity of humans. Nowadays, humanoid robots are able of displaying motor dexterities for dancing, playing musical instruments, talking, etc. Although the long-term goal of true autonomous humanoid robots has yet to be accomplished, the

DOI: 10.4018/978-1-4666-1945-6.ch034

feasibility of integrating them into people's daily lives is becoming closer.

Towards developing humanoid robots capable of interacting more naturally with human partners, robots are required to process and display human-like emotions. The way a person interacts with a humanoid robot is quite different from interacting with the majority of industrial robots today. Modern robots are generally viewed as tools that human specialists use to perform hazardous tasks in remote environments. In contrast, human-like personal robots are often designed to engage people in order to achieve social or emotional goals. The development of socially intelligent and socially skillful robots drives research to develop autonomous or semi-autonomous robots that are natural and intuitive for the average consumer to interact with, communicate with, work with as partners, and teach new capabilities. In addition, this domain motivates new questions for robotics researchers, such as how to design for a successful long-term relationship where the robot remains appealing and provides consistent benefit to people over weeks, months, and even years. The benefit that social robots provide people extends far beyond the strict task performing utility to include educational, health and therapeutic, domestic, social and emotional goals (e.g., entertainment, companionship, communication, etc.), and more.

However, these mechanical devices are still far from understanding and processing emotional states as humans do. Research on musical performance robots seems like a particularly promising path toward helping to overcome this limitation, because music is a universal communication medium, at least within a giving cultural context. Furthermore, research into robotic musical performance can shed light on aspects of expression that traditionally have been hidden behind the rubric of "musical intuition." The late Prof. Ichiro Kato argued that the artistic activity such as playing a keyboard instrument would require human-like intelligence and dexterity (Kato, et al., 1973). In 1984, at Waseda University, the WABOT-2 was the first attempt of developing an anthropomorphic music robot capable of playing a concert organ (Sugano & Kato, 1987). Then, in 1985, the WASUBOT built also at Waseda, could read a musical score and play a repertoire of 16 tunes on a keyboard instrument. More recently, thanks to the technological advances on power computation, Musical Information Retrieval (MIR) and Robot Technology, several researchers have been focusing on developing anthropomorphic robots and interactive automated instruments capable of interacting with musical partners. As a result, different kinds of wind playing-instrument automated machines and humanoid robots have been developed for playing wind instruments (Doyon & Liaigre, 1966; Klaedefabrik, 2005; Solis, et al., 2008; Takashima & Miyawaki, 2006; Solis, et al., 2009a; Dannenberg, 2005; Toyota Motor Corporation, 2011; Degallier, 2006; etc.). Other researchers have been focusing in analyzing wind instrument playing from a musical engineering approach by performing experiments with simplified mechanisms (Ando, 1970; Guillemain, et al., 2010; etc.) and from a physiological point of view by analyzing medical imaging data of professional players (Mukai, 1992; Fletcher, 2001; etc.). In this research, we particularly deal with the development of an anthropomorphic saxophone-playing robot designed to mechanically emulate the required organs during the saxophone playing. Due to the interdisciplinary nature of this research, our collaboration with musicians, musical engineers, and medical doctors will certainly contributes to better reproduce and understand the human motor control from an engineering point of view.

Certainly, the performance of any musical instrument is not well defined and far from a straightforward challenge due to the many different perspectives and subject areas. An idealized musical robot requires many different complex systems to work together integrating musical representation, techniques, expressions, detailed control and sensitive multimodal interactions within the context of a piece, as well as interactions between performers and the list grows. Due to the inherent interdisciplinary nature of the topic, this research can contribute to the further enhance musical understanding, interpretation, performance, education, and enjoyment.

However, if we consider the use of such a complex mechanisms to introduce undergraduate students the principles of robot technology, there could be difficulties to have experience on-hands with an anthropomorphic robot. On the other hand, the continuous falling of the birthrate in developed countries is resulting in a reduction in the number of students where most of them are going away from scientific fields. This situation may tremendously affect the industry by losing competitive power in the future due to the shortage of talented engineers. Moreover, the curricula of engineering universities is currently lacking in practical, design elements resulting in a shortage of opportunities for promoting the creativity of students. For this purpose, several attempts to built educational robots have been done during the past few decades (Miller, et al., 2008).

DEVELOPMENT OF ANTHROPOMORPHIC MUSICAL ROBOTS

Background

During the golden era of automata, the "Flute Player" developed by Jacques de Vaucanson was designed and constructed as a means to understand the human breathing mechanism (Doyon & Liaigre, 1966). Vaucanson presented "The Flute Player" to the Academy of Science in 1738. For this occasion, he wrote a lengthy report carefully describing how his flutist can play exactly like a human. The design principle was that every single mechanism corresponded to every muscle (Vaucanson, 1979). Thus, Vaucanson had arrived at those sounds by mimicking the very means by which a man would make them. Nine bellows were attached to three separate pipes that led into the chest of the figure. Each set of three bellows was attached to a different weight to give out varying degrees of air, and then all pipes joined into a single one, equivalent to a trachea, continuing up through the throat, and widening to form the cavity of the mouth. The lips, which bore upon the hole of the flute, could open and close; and move backwards or forwards. Inside the mouth was a moveable metal tongue, which governed the airflow and created pauses.

More recently, the "Flute Playing Machine" developed by Martin Riches was designed to play a specially made flute somewhat in the manner of a pianola, except that all the working parts are clearly visible (Klaedefabrik, 2005). The Flute Playing Machine is composed of an alto flute, blower, electro-magnets, and electronics. The design principle is basically transparent in a double sense. The visual scores can be easily followed so that the visual and acoustic information is synchronized. The pieces it plays are drawn with a felt tip pen on long transparent music rolls, which are then optically scanned by the photo cells of a reading device. The machine has a row of 15 photocells, which read felt-tip pen markings on a transparent roll. Their amplified signals operate the 12 keys of the flute and the valve, which controls the flow of air into the embouchure. The two remaining tracks may be used for regulating the dynamics or sending timing signals to a live performer when performing a duet.

Since 1990, the authors have been focusing on development of the anthropomorphic flutist robot designed to mechanically emulate the anatomy and physiology of the organs involved during flute playing. In 2007, the Waseda Flutist Robot No. 4 Refined IV (WF-4RIV) was developed. The WF-4RIV has a total of 41-DOFs and it is composed of the following simulated organs (Solis, et al., 2008): lungs, lips, tongue, vocal cord, fingers, and other simulated organs to hold the flute (i.e. neck and arms). The lips mechanism is composed by 3-DOFs to realize an accurate control of the motion of the superior lip (control of airstream's thickness), inferior lip (control of airstream's angle) and sideway lips (control of airstream's length). The artificial lip is made of a thermoplastic rubber named "Septon" (Kuraray Co. Ltd., Japan). The lung system is composed of two acrylic cases, which are sealed. Each of the cases contains a bellow, which is connected to an independent crank mechanism. The crank mechanism is controlled by using an AC motor so that the robot can breathe air into the acrylic cases and breathe air out from them by controlling the speed of motion of the bellow. Finally, the vocal cord is composed by 1-DOF and the artificial glottis is also made of Septon. In order to add vibration to the incoming air stream, a DC motor linked to a couple of gears is used

One of the first attempts to develop a saxophone-playing robot was done by Takashima at Hosei University (Takashima & Miyawaki, 2006). Such a robot, named APR-SX2, is composed of three main components: mouth mechanism (as a pressure controlled oscillating valve), the air supply mechanism (as a source of energy), and fingers (to make the column of air in the instrument shorter or longer). The artificial mouth consisted of flexible artificial lips and a reed pressing mechanism. The artificial lips were made of a rubber balloon filled with silicon oil with the proper viscosity. The air supplying system (lungs) consists of an air pump and a diffuser tank with a pressure control system (the supplied air pressure is regulated from 0.0 MPa to 0.02 MPa). The APR-SX2 was designed under the principle that the instrument played by the robot should not be changed. A finger mechanism was designed to play the saxophone's keys (actuated by solenoids), and a modified mouth mechanism was designed to attach it to the mouthpiece, no tonguing mechanism was implemented (normally reproduced by the tongue motion). The control system implemented for the APR-SX2 is composed by one computer dedicated to the control of the key fingering, air pressure and flow, pitch of the tones, tonguing, and pitch bending. In order to synchronize all the performance, the musical data was sent to the control computer through MIDI in real-time. In particular, the SMF format was selected to determine the status of the tongue mechanism (on or off), the vibrato mechanism (pitch or volume), and pitch bend (applied force on the reed).

Hosei University has developed the APR-SX2; its design is based on the concept of reproducing melodies on a tenor saxophone. Therefore, the saxophone playing robot has been developed under the condition that the musical instrument played by robots should not be changed or remodeled at all. However, a total of twenty-three fingers have been used to play the saxophone's keys (actuated by solenoids), a modified mouth mechanism has been designed (composed by a flexible artificial lip and a reed pressing force control mechanism were developed) to attach it with the mouthpiece, and no tonguing mechanism has been implemented (normally reproduced by the tongue motion).

In contrast, authors proposed in Solis et al. (2009b) the development of an anthropomorphic saxophonist robot as an approach to enable the interaction with musical partners. Therefore, as a long-term goal, we expect that the proposed saxophonist robot is able not only of performing a melody, but also to dynamically interact with the musical partner (i.e. walking while playing the instrument, etc.). As a first result of our research, we have presented the Waseda Saxophonist Robot No. 1 (WAS-1), which it was composed by 15 Degrees of Freedom (DOF) required to play an alto saxophone (Solis, et al., 2009a). In particular, lower lip (1-DOF), tongue (1-DOF), oral cavity, artificial lungs (air pump: 1-DOF and air flow valve: 1-DOF), and fingers (11-DOFs) were developed. Both lips and oral cavity were made of a thermoplastic rubber (named Septon and produced by Kuraray Co.). An improved version, the Waseda Saxophonist Robot No. 2 (WAS-2) was presented, where the design of the artificial lips was improved and a human-like hand was designed (Solis, et al., 2010a). Furthermore, an

Overblowing Correction Controller was implemented in order to assure the steady tone during the performance by using the pitch feedback signal to detect the overblowing condition and by defining a recovery position to correct it (Solis, et al., 2010b). However, the range of sound pressure was still too limited to reproduce the dynamic effects of the sound (i.e. decrescendo) and deviations on the pitch were detected. Therefore, the design of the oral cavity shape has been improved to expand the range of sound pressure and potentiometers were attached to each finger for implementing a dead-time compensation controller. From the control system point of view, a Pressure-Pitch Controller has been proposed to ensure the accurate control of the pitch during the steady phase of the sound produced by the saxophone. Thus, in the following sub-section, we describe the mechanical improvements on the oral cavity and finger mechanisms. In addition, the implementation of a finger dead-time compensation controller and Multiple-Input Multiple-Output controller to assure the accurate control of both air pressure and sound pitch.

Anthropomorphic Saxophonist Robot: Mechanism Design and Control Implementation

In 2010, we have developed the Waseda Saxophonist Robot No. 2 Refined (WAS-2R), which has improved the shape of the oral cavity for increasing the sound range volume and added sensors to each finger for reducing the response delay. In particular, the WAS-2R is composed by 22-DOFs that reproduce the physiology and anatomy of the organs involved during the saxophone playing as follows (Figure 1): 3-DOFs to control the shape of the artificial lips, 16-DOFs for the human-like hand, 1-DOF for the tonguing mechanism, and 2-DOFs for the lung system. In addition, to improve the stability of the pitch of the sound produced, a pressure-pitch controller system has been implemented.

In the previous mechanism, it was possible to confirm the enhancement of the sound range produced by WAS-2 (Solis, et al., 2010a). However, we detected that the note C3 was not possible to be produced. Therefore, we considered to analyze in more detail the oral cavity (in particular, the gap between the palate and the tongue) of professional saxophonist while playing the instrument. For this purpose, we have used an ultrasonic sound probe (ALOKA ProSound II, SSD-6500SV) to obtain images of the oral cavity from professional players while producing the sound of the note C4. By analyzing the obtained images, when a higher volume sound is produced, a large gap between the palate and the tongue is observed. In contrast, while producing lower volume sounds, the gab is considerably narrowed. As a result from these measurements, a new oral cavity for the WAS-2R has been designed (Figure 2). Basically, based on the measurements obtained from images obtained from the professional player, the sectional area has been designed with 156 mm^2 (previous one was 523 mm^2).

Figure 1. The Waseda saxophonist robot no. 2 refined (WAS-2R)

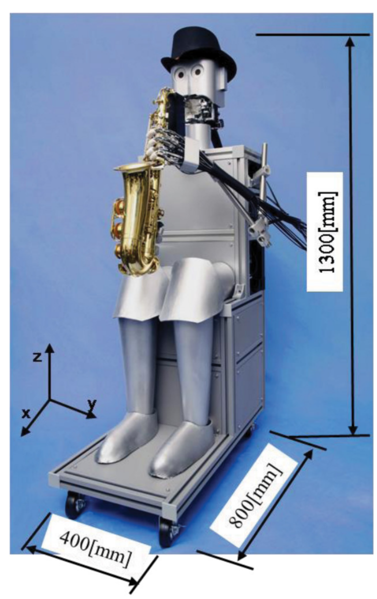
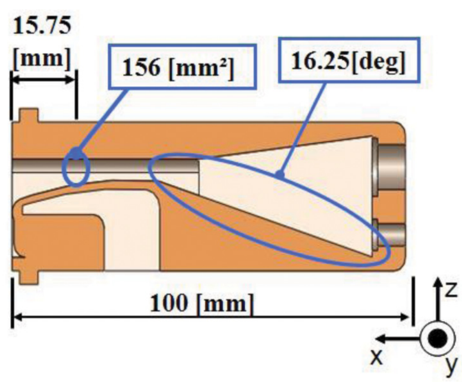


Figure 2. Detail of the oral cavity of WAS-2R



In the previous mechanism, a human-like hand (actuated by a wire-driven mechanism) had been designed to enable the WAS-2 to push all the keys of the alto saxophone (Solis, et al., 2010a). However, due to the use of the wire-driven mechanism, a dynamic response delay (approximately 110ms) has been observed. Therefore, in order to reduce such a delay time, we proposed to embed sensors for measuring the rotational angle of each finger (Figure 3). For this purpose, a rotary sensor (RDC506002A from Alps Co.) has been embedded into the each finger mechanism. In particular, each sensor was placed on a fixing mount device produced by a rapid prototyping device (CONNEX 500). As a result, we were able of attaching the sensing system without increasing the size of the whole mechanism. RC servo motors have been used to control the wire-driven mechanism designed for each finger. As end-effector, an artificial finger made of silicon has been designed. In order to control the sixteen RC motors, the RS-485 serial communication protocol has been used.

On the other hand, the previous mouth mechanism was designed with 1-DOF in order to control the vertical motion of the lower lip. Based on the up/down motion of the lower lip, it became possible to control the pitch of the saxophone sound. However, it is difficult to control the sound pressure by means of 1-DOF. Therefore, the mouth mechanism of the WAS-2 consists of 2-DOFs designed to control the up/down motion of both lower and upper lips (Figure 4a). In addition, a passive 1-DOF has been implemented to modify the shape of the sideway lips. The artificial lips were also made of Septon. In particular, the arrangement configuration of the lip mechanism is as follows: upper lip (rotation of the motor axis is converted into vertical motion by means of a timing belt and ball screw to avoid the leak of air flow), lower lip (a timing belt and ball screw so that the rotational movement of the motor axis is converted into vertical motion to change the amount of pressure on the reed), and sideway lip.

In order to select the motor for the mouth mechanism, the required force for pressing the reed and the maximum stroke of the pins embedded in lip were considered. The target time for

Figure 3. Details of the finger mechanism of WAS-2R

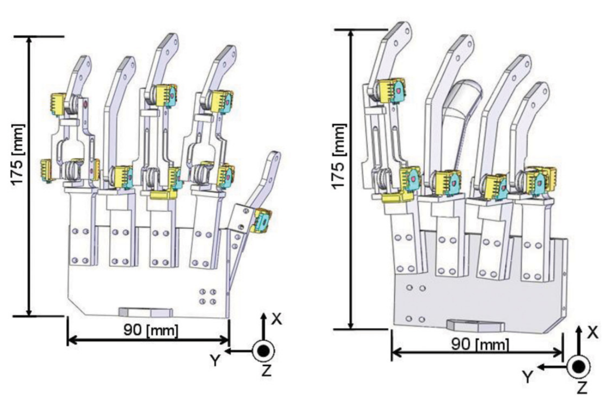
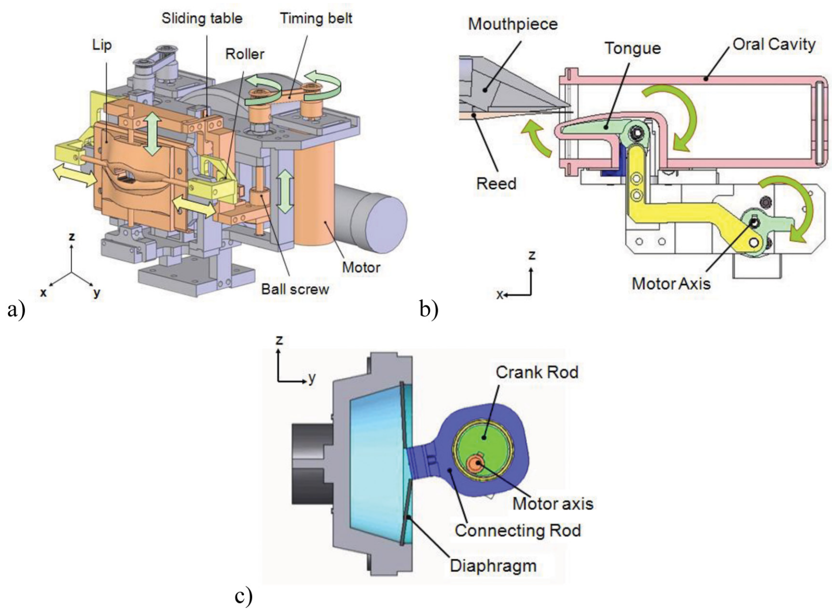


Figure 4. Mechanism details of the WAS-2R: a) mouth mechanism; b) tonguing mechanism; c) lung mechanism



the positioning was set to 100 ms. In order to assure a compact design for the mechanism, a ball screw and timing belt were used. Due to the space constrains, the ball screw SG0602-45R88C3C2Y (KSS Co.) was used. The shaft diameter is 6 mm, and the lead is 2 mm. From those, the axial direction allowable load and allowable revolution were calculated. The requirement of the system is to move 10 mm in 100 ms. Therefore, the average speed v and acceleration a are 0.1 m/s and 4 m/s^2 respectively.

In order to move the pin attached to both sides of lip, the total mass of moving part is 0.05 kg. The axial direction load generated when pin is pulled is given by (1), and this value is the maximum axial direction load applied to the ball screw. The core diameter of the ball screw is 5.1 mm; therefore, the screw shaft minimum moment of inertia of area is given by (2).

$$\begin{aligned} Fa &= 8 + ma \\ &= 8 + 0.05 \times 4 \\ &= 8.2 \quad [N] \end{aligned} \qquad (1)$$

$$I = \frac{\pi}{64} d_1^4 = \frac{\pi \times 5.1^4}{64} = 33.2 \quad [mm^4] \qquad (2)$$

The buckling load is computed by (3), where l_a is the distance between two mounting surfaces (40 mm), E the Young's modulus (2.1×10^5 N/mm^2) and η_1 the factor according to the mounting method (2.0). As a result of the above calculations, we confirmed that the selected ball screw is safe in use.

$$P_1 = \frac{\eta_1 \cdot \pi^2 \cdot E \cdot I}{l_a^2} \times 0.5 = 12485.6 \quad [N] \qquad (3)$$

Then, we have verified the critical speed. Due to the reduction ratio is 1, the required motor revolution is given by (4); the sectional area S of the screw axis is computed by (5).

$$Nm = \frac{V_{\max} \cdot 1000 \cdot 60}{l} \times \frac{1}{A} = \frac{0.20 \times 1000 \times 60}{2} \times \frac{1}{1} = 6000 \quad [rpm] \tag{4}$$

$$S = \frac{\pi}{4} \times 5.1^2 = 20.4 \quad [mm^2] \tag{5}$$

Finally, the allowable revolution of threaded shaft can be computed as (6), where S is the section area (20.4 mm^2), γ the density (7.85×10^{-6} kg/mm^3) and λ_1 is the factor according to the mounting method (3.927). From the above calculations, we could confirm the required revolution is allowable. Thus, we decided to use this ball screw.

$$N_1 = \frac{60 \cdot \lambda_1^2}{2\pi \cdot l_a^2} \times \sqrt{\frac{E \times I \times g}{\gamma \cdot S}} \times 0.8 = \frac{60 \times 3.927^2}{2\pi \times 40^2} \times \sqrt{\frac{2.1 \times 10^5 \times 6.21 \times 9.8 \times 10^3}{7.85 \times 10^{-6} \times 20.4}} \times 0.8 = 1520061 \quad [rpm] \tag{6}$$

After confirming the ball screw specifications, the selection of the motor was verified. For the mouth mechanism, the motor RE-25 (Maxon Co.) was used. In order to calculate the rotary torque required to translate rotary motion into linear motion, the required rotary torque T_1 for an external load is defined as (7), where η is the efficiency of ball screw (0.9).

$$T_1 = \frac{Fa \cdot l}{2\pi \cdot \eta} \cdot A = \frac{10.2 \times 2}{2\pi \times 0.9} \times 1 = 2.90 [N \cdot mm] \tag{7}$$

Because the preload torque T_d of the selected ball screw is 3.0-7.0 Nmm, the preload torque generated T_2 is defined as (8).

$$T_2 = T_d \cdot A = 7.0 \times 1 = 7.0 \; [N \cdot mm] \tag{8}$$

Considering inertia moment of screw shaft and the pulley on the side of motor, the inertia moment J is computed as (9); where J_S is the Inertia moment of screw shaft (2.5×10^{-8} kg*m^2) and J_B is the Inertia moment of pulley on the side of motor (9.11×10^{-7} kg*m^2).

$$J = m\left(\frac{l}{2\pi}\right)^2 \times 10^{-6} + J_S + J_B = 0.05 \times \left(\frac{2}{2\pi}\right)^2 \times 10^{-6} + 2.50 \times 10^{-8} + 9.11 \times 10^{-7} = 9.41 \times 10^{-7} \; [kg \cdot m^2] \tag{9}$$

Because the acceleration time is 0.05 sec, the angular acceleration is computed as (10). Therefore, the required acceleration torque T_3 is given by (11).

$$\dot{\omega} = \frac{2\pi \cdot N_m}{60t} = \frac{2\pi \times 6000}{60 \times 0.050} = 12566.3 \; [rad / \sec^2] \tag{10}$$

$$T_3 = J \times \dot{\omega} \times 10^3 = 9.41 \times 10^{-7} \times 12566.3 \times 10^3 = 11.83 \; [N \cdot mm] \tag{11}$$

From the torques calculated above, the total required acceleration torque T_k is given by (12). The effective value of torque required to the motor is then computed as (13). As a result from the calculations below, it is verified that the RE-25 motor covers the required specifications.

$$T_K = T_1 + T_2 + T_3 = 2.90 + 7.0 + 11.83 = 21.73 \; [N \cdot mm] \tag{12}$$

$$T_{rms} = \sqrt{\frac{T_1^2 \times t_1 + T_2^2 \times t_2 + T_3^2 \times t_3}{t}} = \sqrt{\frac{2.90^2 \times 0.10 + 7.0^2 \times 0.10 + 21.73^2 \times 0.05}{0.10 + 0.10 + 0.05}} = 5.033\ [N \cdot mm] \quad (13)$$

On the other hand, the tonguing mechanism is shown in Figure 4b. The motion of the tongue tip is controlled by a DC motor which is connected to a link attached to the motor axis. In such a way, the airflow can be blocked by controlling the motion of the tongue tip. Thanks to this tonguing mechanism of the WAS-2, the attack and release of the note can be reproduced. In order to select the motor for tongue mechanism, we assumed a response time of 20 ms. As the motor of the tongue mechanism should rotate 20deg in 20ms, the average angular speed is 17.45 rad/s. On the other hand, to approximate the real lingual motion speed, the maximum angular speed is 34.9 rad/s. Therefore, acceleration of it is 3490.7 rad/s^2. The required torque to rotate the tongue mechanism covered with SEPTON is 5.5×10^{-2} Nm and the inertia moment of the center of rotation generated to a part rotate with tongue is 1.19×10^{-5} kg*m^2. Therefore, the required total torque T_{total} for driving the tongue mechanism is computed by (14).

$$T_{total} = T + I\ddot{\theta} = 0.09654[N \cdot m] = 96.54[N \cdot mm] \quad (14)$$

As a result of calculations above, and because of motor size, the motor RE-30 (Maxon Co.) is selected for tongue mechanism.

Regarding the WAS-2R's air source, a DC servo motor has been used to control the motion of the air pump diaphragm, which is connected to an eccentric crank mechanism (Figure 4c). This mechanism has been designed to provide a minimum 20 L/min airflow and a minimum pressure of 30kPa. In addition, a DC servo motor has been designed to control the motion of an air valve so that the delivered air by the air pump is effectively rectified. In order to select the motor for the lung mechanism, the requirement specification was based on the maximum oral cavity pressure (8 kPa) and the calculations of the external force F computed by (15), where F_a is the inertia, F_k is the spring and F_p is the pressure. The force applied to motor arm F_l is then computed by (16), where θ is the angle of rotation and ϕ is the angle of arm. Finally, based on the motor load torque T given by (17), where r is the arm length, the motor RE-30 (Maxon Co.) has been selected.

$$F = Fa + Fk + Fp \quad (15)$$

$$F_l = \frac{F}{\cos\phi} \cdot \sin(\phi + \theta) \quad (16)$$

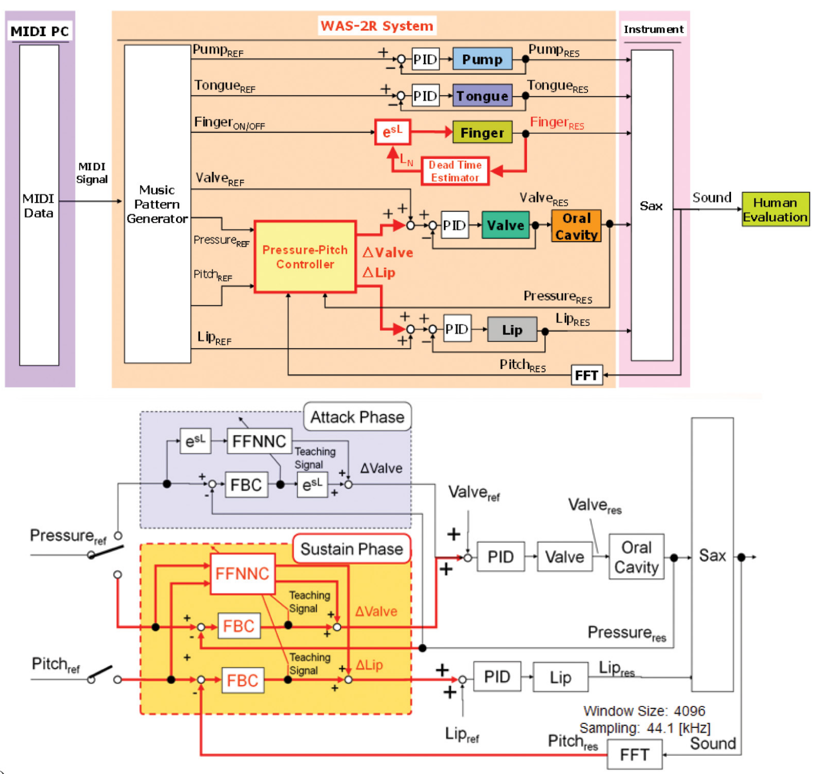
$$T = F_l \cdot r \quad (17)$$

Regarding the control system in our previous research, a feed-forward air pressure controller with dead-time compensation has been implemented to ensure the accurate control of the air pressure during the attack time (Solis, et al., 2010b). Moreover, for the control of the finger mechanism, a simple ON/OFF controller has been implemented. In particular, the feedback error learning during the attack phase of the sound has been used to create the inverse dynamics model of the Multiple-Input Single-Output (MISO) controlled system based on Artificial Neural Networks (ANN). In addition, an Overblowing Correction Controller (OCC) has been proposed and implemented in order to ensure the steady tone during the performance by using the pitch feedback signal to detect the overblowing condition and by defining a recovery position (off-line) to correct it (Solis, et al., 2010b). However, we still detect deviations on the pitch while playing the saxophone.

Therefore, we proposed the implementation of the control system shown in Figure 5a. In particular, the improved control system includes a dead-time compensation controller for the finger mechanism (to reduce the effect of response delay due to the wire-driven mechanism) and a Pressure-Pitch Controller (PPC) for the control of the valve and lip mechanism (to assure the accurate control of the pitch). Regarding the implementation of the dead-time compensation control; for each finger of WAS-2R, the pressing time of the saxophone's key is measured by means of the embedded potentiometer sensor (defined as L_N; where N represents the total number of DOFs designed for the finger mechanism). By including the dead-time factor (referred as e^{sL}), it is possible to compensate the finger's response delay during the saxophone playing (Kim, et al., 2003).

As for the implementation of the control system, a pressure-pitch controller during the sustain phase of the sound has been proposed not only to ensure the accurate control of the air pressure during the attack phase of the sound, but also to ensure the accurate control of both air pressure and sound pitch during the sustain phase of the sound. For this purpose, we implemented a feed-forward error learning method (Kawato & Gomi, 1992) to create the inverse model of the proposed Multiple-Input-Multiple-Output (MIMO) system which is computed by means of an ANN. During the training process, the inputs of the ANN are defined as follows (Figure 5b): pressure reference ($Pressure_{REF}$), pitch reference ($Pitch_{REF}$). In this

Figure 5. Detail of the control system implemented for the WAS-2R: a) block diagram of the improved control system; b) detail of the ANN during the learning phase based on the feedback error learning method



a)

b)

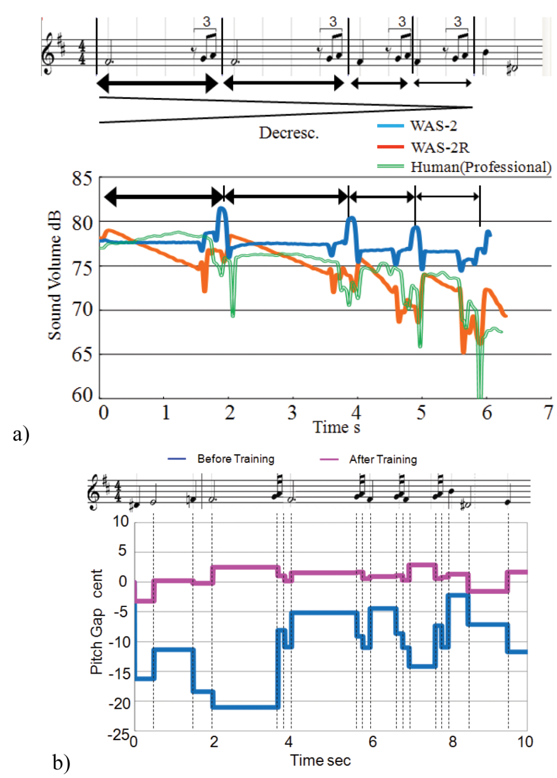
case, a total of six hidden units were used (experimentally determined while varying the number of hidden units). As an output, the position of the air valve (*ΔValve*) and lower lip (*ΔLip*) are controller to ensure the accurate control of the required air pressure and pitch to produce the saxophone sound. Moreover, during the training phase, the air pressure ($Pressure_{RES}$) and sound pitch ($Pitch_{RES}$) are used as feedback signals and both outputs from the feedback controller are used as teaching signals for the effectively training the ANN. As a result from the training phase, during a saxophone playing performance, the created inverse model is used.

Musical Performance

In order to verify if the re-designed shape of the oral cavity contributes to extend the range of sound pressure, we have compared the previous mechanism with the new one while playing the notes from C3 to C5. The average sound pressure ranges for WAS-2R and WAS-2 are 17.7 dB and 9.69 dB, respectively. Moreover, an intermediate player and professional are 13.2 and 22.6 respectively. From this result, we confirmed an increment of 83% thanks to the new shape of the oral cavity. Therefore, we could conclude that the shape of the gap between the palate and tongue has a big influence on the sound pressure range. Thanks to this considerable improvement on the range of sound pressure, we proposed to compare the reproduction of the decrescendo, which is a dynamic sound effect that gradually reduces the loudness of the sound. For this purpose, we programmed the WAS-2 and WAS-2R to play the principal theme of the "Moonlight Serenade" composed by Glenn Miller. The experimental results are shown in shown in Figure 6a. As we may observe, the WAS-2R was able of reproducing nearly similar to the performance of the professional one.

On the other hand, in order to determine the effectiveness of the proposed pressure-pitch controller to reduce the pitch deviations while playing the saxophone, we programmed the WAS-2R to play the main theme of the "Moonlight Serenade" composed by Glenn Miller before and after training the inverse model. In particular, as for the neural network parameters, a total of 6 hidden units were used. For the training process, a total of 144 steps were done. The experimental results are shown in Figure 6b; where 1[cent] is defined as (equi-tempered semitone/100). As we could observe, the deviations of the pitch after the training (Standard Error is 41.7) are considerable less than before training (Standard Error is 2372.8).

Figure 6. Experimental results: a) reproduction of decrescendo effect; b) comparing the deviations of the pitch before and after training the inverse model of the proposed MIMO system with the WAS-2R.



DEVELOPMENT OF EDUCATIONAL ROBOTS

Background

Even though several universities and companies have been building robotic platforms for educational purposes, we may observe that there is still no platform designed to intuitively introduce the principles of RT from the fundamentals to their application to solve real world problems. In fact, most of the current educational platforms focus on providing the basic components to enable students building their own designed system. However; such kind of platforms are used to merely introduce basic control methods (i.e. Sequential Control), basic programming (i.e. Flow Chart Design, C language), and basic mechanism design.

As an approach to cover different aspects of the Robot Technology, in this project we focused in developing an education tool designed to introduce at different educational levels the principle of developing mechatronic systems. In particular, the development of an inverted pendulum mobile robot has been proposed. In fact, the inverted pendulum has been the subject of numerous studies in automatic control (Grasser, et al., 2002; Salerno & Angeles, 2007; Koyanagi, et al., 1992; Kim, et al., 2003; Pathak, et al., 2005; etc.), introduction to Mechatronics (Solis & Takanishi, 2009; etc.), etc.

Up to now, several attempts to build educational robots have been made during the past few decades (Miller, et al., 2008). In fact, the development of educational robots started in the early 1980s with the introduction of the Heathkit *Hero − 1* (Heath Co.). Such kind of robot was designed to encourage students to learn how robots are built. However, no information on the theory or principles behind the assembly is given. More recently, several other companies in cooperation with universities and research centers have been trying to introduce educational robots to the market. Some examples are as follows: K-Team (K-TEAM Ltd.) introduces the *Hemisson*, which is a low-cost educational robot designed to provide an introduction to robot programming by using reduced computational power and few sensors. Another example is the *LEGO ® Mindstorms RCX,* which is a good tool for early and fast robot design by using the LEGO blocks (LEGO Ltd.). In Japan, we can also find some examples such as: the *RoboDesigner* kit designed to provide a general platform to enable students to build their own robots (Japan Robotech Ltd.), *ROBOVIE − MS* from ATRRobotics designed as an education tool to introduce principles of mechanical manufacturing, assembly, and operational programming of small-sized humanoid robot, etc.

From the perspective of introducing RT technology to undergraduate students, it is a good example to provide experience to them on control designing, signal processing, distributed control systems and the consideration of real-time constraints for real applications purposes. However, most of the current proposed robots do not consider the educational issues while designing the inverted pendulum (i.e. possibility of changing the center of mass, etc.). In addition, authors consider the importance to consider the introduction of human-robot interaction to motivate their further interest (i.e. the size of the robot should fit the size of a personal mobile computer, etc.).

Therefore, the authors have proposed the development of a two-wheeled inverted pendulum type mobile robot designed to cover the basic principles in electronics, mechanical engineering, programming, as well as, more advanced topics on control engineering, complex programming, and embedded systems. As a result of our research, the Waseda Wheeled Vehicle No.2 Refined (WV-2R) has been introduced (Solis, et al., 2009c). In particular, the WV-2R has been designed to enable students to verify the changes on the response of the robot while varying some physical parameters of the robot. From the experimental results, we confirm some of the educational functions of the proposed robot (i.e. PID tuning, varying the center of mass, etc.). However, a hand-made control board has been used so that several problems of wire connections were detected. Furthermore, the WV-2R didn't include any additional mechanism for proposing different kinds of robot contest. Finally, from our discussions with undergraduate students, the development of a simulator could considerably increase their knowledge.

Two-Inverted Pendulum Robot: Mechanism Design and Control Implementation

In the 2010, the Waseda Wheeled Vehicle Robot No. 2 Refined II (WV-2RII) was developed as an educational robot designed to implement different educational issues to introduce undergraduate students the principles of RT (Figure 7). The specifications are shown in Table 1. The WV-2RII is composed of two-actuated wheels, a general-purpose control board (Figure 8a), an adjustable weighting bar attached to the pendulum, a gyro and accelerometer sensors, a remote controller (Figure 8b), and two optional mechanisms that can be easily attached/detached from the main body of the robot. In particular, the general-purpose control board consists of a 32 bits ARM microcontroller, 10 general I/O ports, 2 motor drives, a LCD display, 8 LEDs, a Zigbee module, and 2 servo connectors.

The WV-2RII is endowed with two active wheels actuated by DC motors. The model description is shown in Figure 9; where the following parameters are defined as follows.

θ: Tilt angle of the chassis
φ: Axial component of the angular velocity of the wheel
m_1: Mass of the chassis
m_2: Wheel mass
J_1: Moment of inertia of the chassis
J_2: Wheel Moment of Inertia
l: Distance between wheel axis and robot mass center
r: Wheel radius

Figure 7. The Waseda wheeled vehicle no. 2 refined II (WV-2RII).

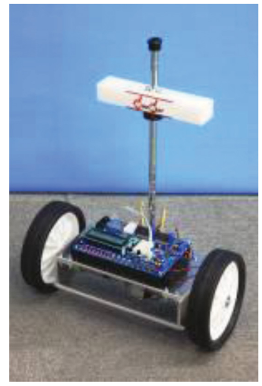


Table 1. The specification of WV-2RII

Parameter	Specifications
Height [mm]	530
Weight [kg]	3.8
DOFs	2-DOFs
Microcontroller	STM32F103VB x 1
Sensors	Accelerometer x 1
	Red Gyro x 1
	Optical Encoder x 2
Motor	RDO-37BE50G9 (12 Volts) x 2
Power Supply	Battery: 6 [V] x 1 RC–Battery: 12 [V] x 1
Remote Controller	ZigBee: 2.6GHz

Figure 8. a) general-purpose control board; b) remote controller for the WV-2RII

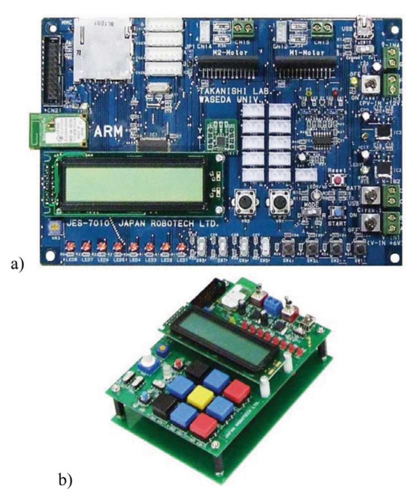
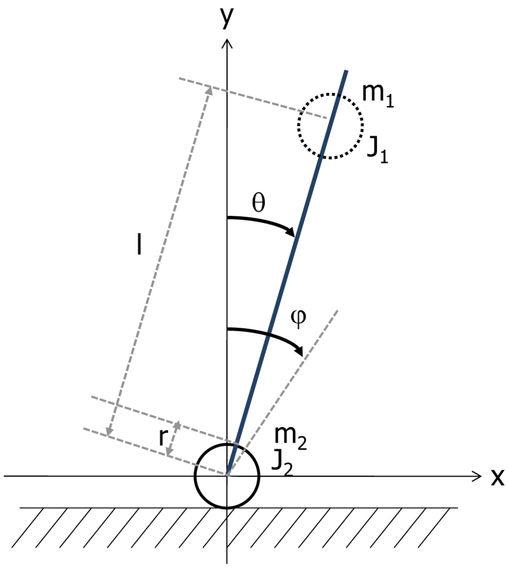


Figure 9. The model of the two-wheeled inverted pendulum robot



By using the above parameters, and by defining T as torque, n as reduction ratio of the gear and S as the Frictional Force on the wheel along the horizontal ground plane (where f_x and f_y are the components of the force acting between the wheel and pendulum at the center of the wheel), we may define the following Equations (18-23):

$$m_2 x_2'' = S - f_x \tag{18}$$

$$J_2 \varphi'' = nT - rS \tag{19}$$

$$m_1 (x_2'' + (l \sin\theta)'') = f_x \tag{20}$$

$$m_1 (l \cos\theta)'' = -m_1 g + f_y \tag{21}$$

$$J_1 \theta'' = f_y l \sin\theta - f_x l \cos\theta - nT \tag{22}$$

$$x = r\varphi \tag{23}$$

Equations (24) and (25) follow from above equations upon elimination of intermediate variables f_x, f_y, S. From Equation (24), we may notice that when the angular acceleration of the body is less than zero, it is possible to correct the vertical inclination of the body to the standing upright position.

$$\varphi'' = \frac{1}{J_2 + r^2 m_1 + r^2 m_2}\left\{nT - m_1 rl(\theta'' \cos\theta - \theta'^2 \sin\theta)\right\} \tag{25}$$

If we define the maximum tilt angle of the chassis to 50 degrees, and use the respective physical parameters corresponding to the WV-2RII (m_1 = 2.247 kg; m_2 = 0.800 kg; J_1 = 0.015 kgm^2; J_2 = 0.002 kg*m^2; g = 9.81 m/s^2; l = 0.0477 m; r = 0.0725 m) into Eq. (28), we obtain the following relation:

$$\therefore nT \geq 2.2[Nm]$$

Based on the above relation, we have selected the motor RDO-37BE50G9 (stall torque 0.160 Nm and Gear ratio 9:1). If we consider the coef-

Equation 24

$$\theta'' = \frac{1}{J_1 + m_1 l^2 + \dfrac{m_1^2 r l^2 \cos^2\theta}{\dfrac{J_2}{r} + rm_1 + rm_2}}\left[m_1 g l \sin\theta - m_1 l \cos\theta \left\{\frac{nT + m_1 rl\theta'^2 \sin\theta}{\dfrac{J_2}{r} + rm_1 + rm_2}\right\} - nT\right]$$

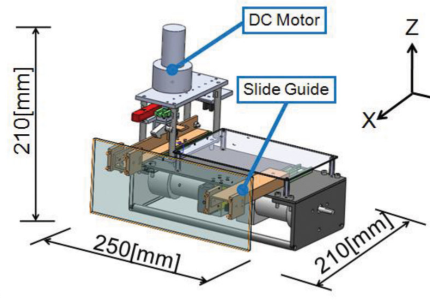
ficient of safety of the power generated by the two motors as 0.8, then nT is 2.3 Nm satisfies the required specification.

On the other hand, as we have previously introduced, we have developed two additional mechanisms that can be easily attached to the main body of the WV-2RII. In particular, a kicking mechanism for soccer (Figure 10a) and an arm mechanism for sumo (Figure 10b) have been designed and constructed. In particular, the soccer-kicking mechanism is composed by a spring, hook, stopper, and a DC motor. In order to kick the ball, a tension spring is used to increase the speed of movement of the kicking mechanism (maximum output load of 22N). Basically, the kicking mechanism is attached to a hook which is moved until a certain point when the hook is automatically released (by a stopper), the reaction force accumulated by the spring is used to kick the ball. On the other hand, the sumo-arm mechanism is composed by sliding-crank mechanism actuated by a DC motor, an arm base actuated by a RC motor to adjust the pitch of the whole arm mechanism and a pushing plate with embedded switches for detecting the contact with the opponent. Basically, in gear wheels of the slider of the crank mechanism, the fixed and movable racks are used. The rotation motion of the crank is transmitted to the gear wheels and the movable rack moves at twice the stroke of the fixed rack. From this, the arm mechanism provides a large stroke (around 88mm) by using a compact mechanism.

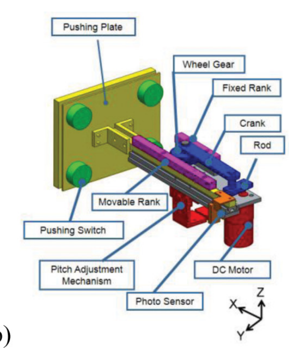
As a further example of application of WV-2RII for showing the potentialities of the proposed system, a female undergraduate student (from mechanical engineering background) during an internship at Waseda University was asked to design an upper body with appearance and gestures that are appealing to children using this new additional robot. For this purpose, we asked the student to design of the upper body mechanism, to develop the required commands for controlling it from a remote controller integrated on the WV-2RII. The detail of the mechanism designed by the internship student is shown in Figure 11a. The proposed upper body uses 4 RC motors to control the motion of head (2-DOFs) and arms/wings (2-DOFs), lending more expression to the robot. Moreover, the possible motions realizable by the upper body are shown in Figure 11b.

In Figure 12, the block diagram of the control system implemented for the WV-2RII is shown. As we may observe, the WV-2RII is controlled by feedback control system. In particular, the rate gyro sensor signal measures the body angular velocity (θ') and the encoder measures the wheel rotational angle (φ). Because the drift on the signal obtained from the gyro is extremely small, the use of a high-pass filer is not required. Therefore, a low-pass filter is only used to compute the

Figure 10. Detail of the additional mechanisms designed for WV-2RII: a) soccer-kicking mechanism; b) sumo-arm mechanism

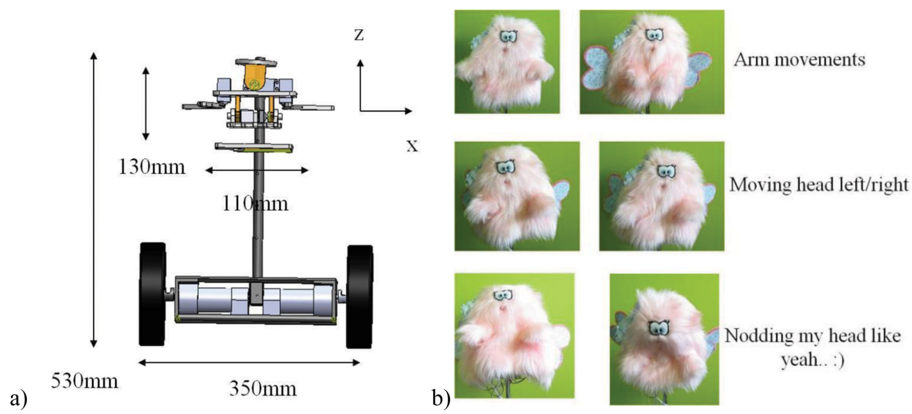


a)



b)

Figure 11. Pictures of the possible motions that the upper body mounted on the WV-2RII

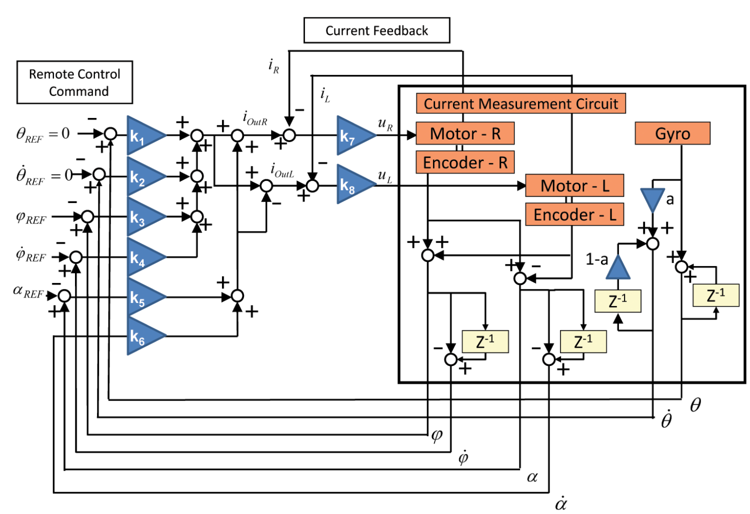


body angular velocity (θ'); where the cut-off frequency is 0.32 Hz. In order to compute the body angle, the body angle and the wheel angular velocity, the body angular velocity and wheel angle are integrated and derivated respectively. In order to control all the parameter, a feedback controller has been implemented by using Equation (26), where k_1~k_6 parameters are the gain coefficients of the controller which are tuned to assure the stabilization of the system. Furthermore, a current feedback controller has been implemented by Equation (27), where the parameter k_7 is tuned for assuring the accurate control of the command current to each motor. As for the command control signal, the θ_{REF}, φ'_{REF}, α'_{REF} are set to zero, while the other commands are sent by a remote controller.

Control Stability

In order to verify the robustness of the proposed controller implemented for the WV-2RII, we have placed the pendulum horizontally on the ground without activating the control. From this starting position, we have activated the control system and given as control goal the vertical position (90

Figure 12. Control block diagram implemented for the WV-2RII



Equations 26 and 27

$$i_{outR} = k_1 \cdot \theta + k_2 \cdot \theta' + k_3 \cdot \left(\phi - \phi_{REF}\right) + k_4 \cdot \left(\phi' - \phi'_{REF}\right) + k_5 \cdot \left(\alpha - \alpha_{REF}\right) + k_6 \cdot \left(\alpha' - \alpha'_{REF}\right)$$
$$u_R = k_7 \cdot \left(i_{outR} - i_R\right)$$

$$i_{outL} = k_1 \cdot \theta + k_2 \cdot \theta' + k_3 \cdot \left(\phi - \phi_{REF}\right) + k_4 \cdot \left(\phi' - \phi'_{RE}\right) - k_5 \cdot \left(\alpha - \alpha_{REF}\right) - k_6 \cdot \left(\alpha' - \alpha'_{REF}\right)$$
$$u_L = k_8 \cdot \left(i_{outL} - i_L\right)$$

degrees). From this experiment, we may observe the dynamic response of WV-2R by analyzing the body angle θ and the motor current measured. The experimental results are shown in Figure 13. As we may observe, the WV-2RII requires around 0.8 sec to reach the target position, where a maximum of 3A is required (the current circuit has been designed to support a peak current up to 7 Amperes).

FUTURE RESEARCH DIRECTIONS

Conventionally, anthropomorphic musical robots are mainly equipped with sensors that allow them to acquire information about its environment. Based on the anthropomorphic design of humanoid robots, it is therefore important to emulate two of the human's most important perceptual organs: the eyes and the ears. For this purpose, the humanoid robot integrates in its head, vision and aural sensors attached to the sides for stereo-acoustic perception. In the case of a musical interaction, a major part of the typical performance (i.e. Jazz) is based on improvisation. In these parts, musicians take turns in playing solos based on the harmonies and rhythmical structure of the piece. Upon finishing his solo section, one musician will give a visual signal, a motion of the body or his instrument, to designate the next soloist. Toward enabling the multimodal interaction between the musician and musical robots, a Musical-based Interaction System (MbIS) will be integrated on the Waseda Saxophonist robot (Figure 14a). The MbIS has been conceived for enabling the interaction between the musical robot and musicians (Petersen, et al.,

Figure 13. Experimental results while programming the WV-2RII to rise from the ground by analyzing the body angle and the applied motor current

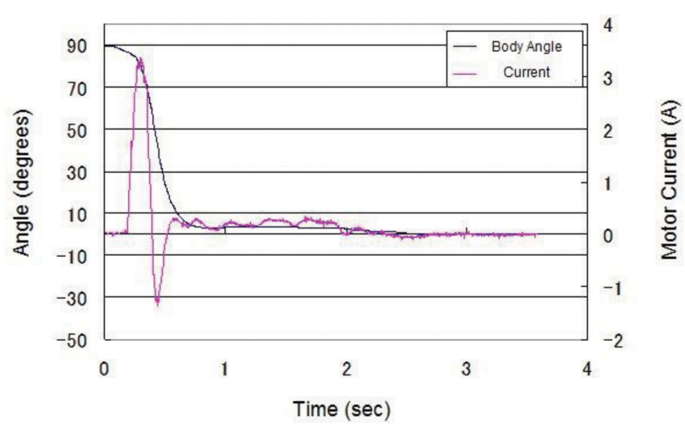
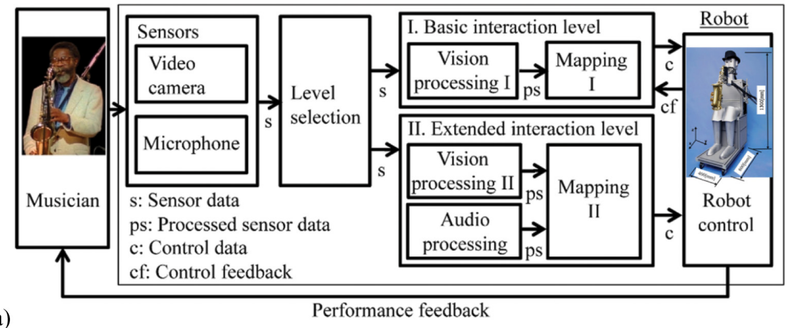
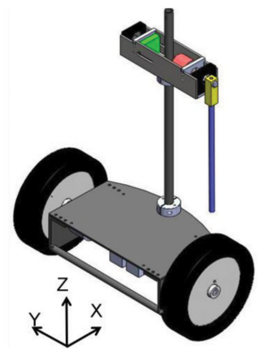


Figure 14. a) proposed musical-based interaction system; b) two-wheeled double inverted pendulum



a)



b)

2010). Even though the WAS-2R still requires several improvements from the mechanical and control point of view, we do expect the robot can be used for the entertainment of elderly people, reproduce the performance of famous saxophonist players passed away and for education of young players as practical applications.

On the other hand, in order to introduce interactive educational robotic systems, the educational platform (both for university students and engineering at the industry) must be designed to cover the basic principles in electronics, mechanics, programming as well as more advanced topics on control, advanced programming and human-robot interaction. Moreover, to enhance the entertainment issue, the educational platform could also include some aspects of art (i.e. music, etc.) to learn other basic aspects such as signal processing (i.e. musical retrieval information, etc), recognition systems (i.e. Hidden Markov Model, etc.), game design (i.e. audio/motion design), etc. Further challenges on dynamic control of a two-wheeled double inverted pendulum robot can be also conceived (Figure 14b). Based in this approach, it is possible to be used in classes beyond the classical Electrical, Mechanics, and Mechatronics Engineering curriculum, including Music Engineering (Martin, et al., 2009; Yanco, et al., 2007), etc. The WV-2RII is now being commercialized as "MiniWay" by Japan Robotech Ltd. Even though this robot has been designed as an educational robot, it is possible to conceive (with some mechanical and control design modifications) different kinds of practical applications such as baggage transportation within an airport, guidance for visitors or entertainment of children at museums, etc.

CONCLUSION

In this chapter, the mechanism design and control implementation proposed for two different human-friendly robotic platforms have been introduced.

In particular, the developments of an anthropomorphic saxophonist robot and a two-wheeled inverted pendulum robot have been detailed. The saxophonist robot has been designed to reproduce the organs involved during the saxophone playing and a feed-forward controller has been implemented in order to accurately control both the air pressure and the sound pitch during a musical performance. On the other hand, the two-wheeled inverted pendulum has been designed to introduce the principles of robot technology at different educational levels and a feedback controller has been implemented in order to assure the stability of the inverted pendulum.

ACKNOWLEDGMENT

Part of the research on the Waseda Saxophonist Robot and Waseda Vehicle Robot Part was done at the Humanoid Robotics Institute (HRI), Waseda University and at the Center for Advanced Biomedical Sciences (TWINs). This research is supported (in part) by a Gifu-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture. This work is also supported (in part) by Global COE Program "Global Robot Academia" from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. Finally, the study on the Waseda Saxophonist Robot is supported (in part) by a Grant-in-Aid for Young Scientists (B) provided by the Japanese Ministry of Education, Culture, Sports, Science, and Technology, No. 23700238 (J. Solis, PI).

REFERENCES

Ando, Y. (1970). Drive conditions of the flute and their influence upon harmonic structure of generated tone. *Journal of the Acoustical Society of Japan,* 297-305.

Dannenberg, R. B., Brown, B., Zeglin, G., & Lupish, R. (2005). McBlare: A robotic bagpipe player. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, (pp. 80-84). ACM.

de Vaucanson, J. (1979). Le mécanisme du fluteur automate: An account of the mechanism of an automation: Or, image playing on the german-flute. In Vester, F. (Ed.), *The Flute Library: First Series No. 5*. Dordrecht, The Netherlands: Uitgeverij Frits Knuf.

Degallier, S., Santos, C. P., Righetti, L., & Ijspeert, A. (2006). Movement generation using dynamical systems: A humanoid robot performing a drumming task. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, (pp. 512-517). IEEE.

Doyon, A., & Liaigre, L. (1966). *Jacques Vaucanson: Mecanicien de genie*. Paris, France: PUF.

Fletcher, N., Hollenberg, L., Smith, J., & Wolfe, J. (2001). The didjeridu and the vocal tract. In *Proceedings of the International Symposium on Musical Acoustics*, (pp. 87-90). ACM.

Grasser, F., D'Arrigo, A., Comlombi, S., & Rufer, A. (2002). Joe: A mobile inverted pendulum. *IEEE Transactions on Electronics*, *49*(1), 107–114. doi:10.1109/41.982254

Guillemain, P., Vergez, C., Ferrand, D., & Farcy, A. (2010). An instrumented saxophone mouthpiece and its use to understand how an experienced musician plays. *Acta Acoustica*, *96*(4), 622–634. doi:10.3813/AAA.918317

Heath Company. (2011). *Website.* Retrieved from http://www.hero-1.com/broadband/.

Japan Robotech Ltd. (2011). *Website.* Retrieved from http://www.japan-robotech.com/eng/index.html.

K-TEAM. (2011). *Website.* Retrieved from http://www.k-team.com.

Kato, I., Ohteru, S., Kobayashi, H., Shirai, K., & Uchiyama, A. (1973). Information-power machine with senses and limbs. In *Proceedings of the CISM-IFToMM Symposium on Theory and Practice of Robots and Manipulators*, (pp. 12-24). ACM.

Kawato, M., & Gomi, H. (1992). A computational model of four regions of the cerebellum based on feedback-error-learning. *Biological Cybernetics*, *68*, 95–103. doi:10.1007/BF00201431

Kim, H., Kim, K., & Young, M. (2003). On-line dead-time compensation method based on time delay control. *IEEE Transactions on Control Systems Technology*, *11*(2), 279–286. doi:10.1109/TCST.2003.809251

Kim, Y. H., Kim, S. H., & Kwak, Y. K. (2003). Dynamic analysis of a nonholonomic two-wheeled inverted pendulum robot. In *Proceedings of the Eighth International Symposium on Artificial Life and Robotics*, (pp. 415-418). ACM.

Klaedefabrik, K. B. (2005). *Martin riches - Maskinerne / the machines*. Berlin, Germany: Kehrer Verlag.

Koyanagi, E., Lida, S., & Yuta, S. (1992). A wheeled inverse pendulum type self-contained mobile robot and its two-dimensional trajectory control. In *Proceedings of ISMCR*, (pp. 891-898). ISMCR.

Kuraray Co. (2011). *Website.* Retrieved from http://www.kuraray.co.jp/en/.

LEGO. (2011). *Website.* Retrieved from http://mindstorms.lego.com/.

Martin, F., Greher, G., Heines, J., Jeffers, J., Kim, H. J., & Kuhn, S. (2009). Joining computing and the arts at a mid-size university. *Journal of Computing Sciences in Colleges*, *24*(6), 87–94.

Miller, D., Nourbakhsh, I., & Siegwart, R. (2008). Robots for education. In Siciliano, B., & Khatib, O. (Eds.), *Springer Handbook of Robotics* (pp. 1287–1290). Berlin, Germany: Springer. doi:10.1007/978-3-540-30301-5_56

Mukai, S. (1992). Laryngeal movement while playing wind instruments. In *Proceedings of International Symposium on Musical Acoustics*, (pp. 239–241). ACM.

Pathak, K., & Franch, J., Agrawal, & Sunil, K. (2005). Velocity and position control of a wheeled inverted pendulum by partial feedback linearization. *IEEE Transactions on Robotics*, *21*(3), 505–513. doi:10.1109/TRO.2004.840905

Petersen, K., Solis, J., & Takanishi, A. (2010). Musical-based interaction system for the waseda flutist robot: Implementation of the visual tracking interaction module. *Autonomous Robots Journal*, *28*(4), 439–455. doi:10.1007/s10514-010-9180-5

Salerno, A., & Angeles, J. (2007). A new family of two wheeled mobile robot: Modeling and controllability. *IEEE Transactions on Robotics*, *23*(1), 169–173. doi:10.1109/TRO.2006.886277

Solis, J., Ninomiya, T. N., Petersen, K., Takeuchi, M., & Takanishi, A. (2009a). Development of the anthropomorphic saxophonist robot WAS-1: Mechanical Design of the simulated organs and implementation of air pressure. *Advanced Robotics Journal*, *24*, 629–650. doi:10.1163/016918610X493516

Solis, J., Petersen, K., Ninomiya, T., Takeuchi, M., & Takanishi, A. (2009b). Development of anthropomorphic musical performance robots: From understanding the nature of music performance to its application in entertainment robotics. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 2309-2314). IEEE Press.

Solis, J., Petersen, K., Yamamoto, T., Takeuchi, M., Ishikawa, S., Takanishi, A., & Hashimoto, K. (2010a). Design of new mouth and hand mechanisms of the anthropomorphic saxophonist robot and implementation of an air pressure feed-forward control with dead-time compensation. In *Proceedings of the International Conference on Robotics and Automation*, (pp. 42-47). ACM.

Solis, J., Petersen, K., Yamamoto, T., Takeuchi, M., Ishikawa, S., Takanishi, A., & Hashimoto, K. (2010b). Implementation of an overblowing correction controller and the proposal of a quantitative assessment of the sound's pitch for the anthropomorphic saxophonist robot WAS-2. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 1943-1948). IEEE Press.

Solis, J., & Takanishi, A. (2009c). Introducing robot technology to undergraduate students at waseda university. In *Proceedings of the ASME Asia-Pacific Engineering Education Congress*. ASME.

Solis, J., Taniguchi, K., Ninomiya, T., & Takanishi, A. (2008). Understanding the mechanisms of the human motor control by imitating flute playing with the waseda flutist robot WF-4RIV. *Mechanism and Machine Theory*, *44*(3), 527–540. doi:10.1016/j.mechmachtheory.2008.09.002

Sugano, S., & Kato, I. (1987). WABOT-2: Autonomous robot with dexterous finger-arm coordination control in keyboard performance. In *Proceedings of the International Conference on Robotics and Automatio*n, (pp. 90-97). ACM.

Takashima, S., & Miyawaki, T. (2006). Control of an automatic performance robot of saxophone: Performance control using standard MIDI files. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems - Workshop: Musical Performance Robots and Its Applications*, (pp. 30-35). IEEE Press.

Toyota Motor Corporation. (2011). *Website*. Retrieved from http://www.toyota-global.com/innovation/partner_robot/.

Yanco, H. A., Kim, H. J., Martin, F. G., & Silka, L. (2007). Artbotics: Combining art and robotics to broaden participation in computing. In *Proceedings of the AAAI Spring Symposium Robots and Robot Venues*. AAAI.

ADDITIONAL READING

Aggarwal, R., Darzi, A., & Sullivan, P. (2009). Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents. *Ophthalmology*, *116*(2), 257–262. doi:10.1016/j.ophtha.2008.09.038

Billard, A. (2003). Robota: Clever toy and educational tool. *Robotics and Autonomous Systems*, *42*, 259–269. doi:10.1016/S0921-8890(02)00380-9

Hashimoto, S. (2009). Kansei technology and robotics machine with a heart. *Kansei Engineering International*, *8*(1), 11–14.

Hayashi, E., Yamane, M., & Mori, H. (200). Development of moving coil actuator for an automatic piano. *International Journal of Japan Society for Precision Engineering, 28*(2), 164–169.

iRobot Corp. (2005). *iRobot roomba serial command interface (SCI) specification*. Burlington, VT: iRobot.

Ishii, H., Koga, H., Obokawa, Y., Solis, J., Takanishi, A., & Katsumata, A. (2009). Development and experimental evaluation of oral rehabilitation robot that provides maxillofacial massage to patients with oral disorders. *The International Journal of Robotics Research*, *28*(9), 1228–1239. doi:10.1177/0278364909104295

Kajitani, M. (1989). Development of musician robots. *Journal of Robotics and Mechatronics*, *1*(3), 254–255.

Kotosaka, S., & Schaal, S. (2001). Synchronized robot drumming by neural oscillator. *Journal of the Robotics Society of Japan*, *19*, 116–123.

Kusuda, Y. (2008). Toyota's violin-playing robot. *Industrial Robot: An International Journal*, *35*(6), 504–506. doi:10.1108/01439910810909493

Martin, F. G. (2001). *Robotic explorations: A hands-on introduction to engineering*. Upper Saddle River, NJ: Prentice Hall.

Matari, M. J., Koenig, N., & Feil-Seifer, D. (2007). *Materials for enabling hands-on robotics and STEM education. Technical Reporty SS-07–09*. Washington, DC: AAAI.

Mayrose, J., Kesavadas, T., Chugh, K., Dhananjay, J., & Ellis, D. E. (2003). Utilization of virtual reality for endotracheal intubation training. *Resuscitation*, *59*(1), 133–138. doi:10.1016/S0300-9572(03)00179-5

Murphy, R. R. (2000). Using robot competitions to promote intellectual development. *Artificial Intelligence Magazine*, *21*(1), 77–90.

Nakadate, R., Matsunaga, Y., Solis, J., Takanishi, A., Minagawa, E., Sugawara, M., & Niki, K. (2011). Development of a robotic-assisted carotid blood flow measurement system. *Mechanism and Machine Theory Journal*, *46*(8), 1066–1083. doi:10.1016/j.mechmachtheory.2011.03.008

Noh, Y., Segawa, M., Shimomura, A., Ishii, H., Solis, J., Hatake, K., & Takanishi, A. (2008). WKA-1R robot-assisted quantitative assessment of airway management. *International Journal of Computer Assisted Radiology and Surgery*, *3*(6), 543–550. doi:10.1007/s11548-008-0238-1

Rowe, R. (2001). *Machine musicianship*. Cambridge, MA: The MIT Press.

Sobh, T. M., & Wange, B. (2003). Experimental robot musicians. *Journal of Intelligent & Robotic Systems*, *38*, 197–212. doi:10.1023/A:1027319831986

Solis, J., Marcheschi, S., Frisoli, A., Avizzano, C. A., & Bergamasco, M. (2007). Reactive robots system: An active human/robot interaction for transferring skill from robot to unskilled persons. *International Advanced Robotics Journal*, *21*(3), 267–291. doi:10.1163/156855307780131992

Solis, J., Oshima, N., Ishii, H., Matsuoka, N., Hatake, K., & Takanishi, A. (2008). Towards an understanding of the suture/ligature skills during the training process by using the WKS-2RII. *International Journal of Computer Assisted Radiology and Surgery*, *3*(3-4), 231–239. doi:10.1007/s11548-008-0220-y

Solis, J., & Takanishi, A. (2010). Recent trends in humanoid robotics research: Scientific background, applications and implications. *Accountability in Research*, *17*, 278–298. doi:10.1080/08989621.2010.523673

Weinberg, G., & Driscoll, S. (2006). Toward robotic musicianship. *Computer Music Journal*, *30*(4), 28–45. doi:10.1162/comj.2006.30.4.28

KEY TERMS AND DEFINITIONS

Anthropomorphic: Musical Robots: A robot designed to reproduce the organs involved during the musical instrument playing able of displaying both motor dexterity and intelligence.

Bio-Inspired Robotics: A robot that mechanically emulates or simulates living biological organisms.

Education Robots: A robot used by students composed by low-cost components commonly found on any robotic platform.

Feed-Forward Error Learning: A computational theory of supervised motor learning that can be used as a training method to compute the inverse dynamics model of the controller system.

Human-Friendly Robotics: Research field focus on the development of new methodologies for the design, control and safety operation of robots designed to naturally and intuitively interact, communicate and work with humans as partners.

Humanoid Robots: A robot designed to reproduce the human body in order to interact naturally with human partners within the human environment.

Inverted Pendulum Robot: A robot composed by an inverted pendulum attached to a mobile base equipped with motors that dive it along a horizontal plane.

This work was previously published in Service Robots and Robotics: Design and Application, edited by Marco Ceccarelli, pp. 130-153, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 35
Dual-SIM Phones:
A Disruptive Technology?

Dickinson C. Odikayor
Landmark University, Nigeria

Ikponmwosa Oghogho
Landmark University, Nigeria

Samuel T. Wara
Federal University Abeokuta, Nigeria

Abayomi-Alli Adebayo
Igbinedion University Okada, Nigeria

ABSTRACT

Dual-SIM mobile phones utilize technology that permits the use of two SIMs at a time. The technology permits simultaneous access to the mobile network services. Its disruptive nature is with reference to the mobile phone market in Nigeria and other parts of the world. Earlier market trend was inclination to "newer" and "better" phones, in favour of established single-SIM mobile phone manufacturers like Nokia and Samsung. Introduction of dual-SIM phones mainly manufactured by Chinese mobile phone manufacturing firms propelled user preference for phones acquisition which permits dual and simultaneous access to mobile network. This technology has compelled its adoption by established manufacturing names in order that they may remain competitive. It is a clear case of a disruptive technology, and this chapter focuses on it need, effects, and disruptive nature.

1.0 INTRODUCTION

Christensen (1997) used the term "*disruptive technology*" in his book *The Innovator's Dilemma*. Such technologies surprise the market by generating a considerable improvement over existing technology, and this can be attained in a number of ways. This technology may not be as expensive as an existing technology or more complicated in nature but does attract more potential users (www.wisegeek.com). At times it may be expensive and complicated, requiring highly skilled personnel and infrastructure to implement. Two types of technology change have shown different effects on the industry leaders. *Sustained technology* sustains the rate of improvement in a

DOI: 10.4018/978-1-4666-1945-6.ch035

product's performance in the industry. Dominant industry firms are always at the fore developing and adopting such technologies. *Disruptive technology* changes or disrupts the performance path and continually results in the failure of the industry leading firms. Few technologies are basically or essentially disruptive or sustaining in nature. It's the impact created by the strategy or business model that the technology enables that is disruptive (Christensen & Raynor 2003). The advent of Global System for Mobile communication (GSM) resulted in a major communication leap worldwide. Mobile phones actually became an indispensible electronic gadget defining the modern world (Sally, Sebire, Riddington, 2010). Mobile phone manufacturers continue to include different features on their mobile phone products in addition to basic functions of communication. This is with a purpose of sustaining the market for the products. The mobile phone has become a gadget with full range of services. Ranging from basic telephony to business and leisure or entertainment features. However, performance issues with mobile network services furnished further basis for multiple SIM (Subscriber Identity Module) acquisition by users, for improved access. The problems that led to this were initially poor network coverage and poor performance problems of mobile network service providers in the country and later lower call tariff. Mobile phone users acquired phones depending on the number of networks to which they were subscribed and the trend still exists today. An opportunity was created for a product that would satisfy user needs with regard to multiple SIM capabilities.

1.1 History of Mobile Phone

The history of mobile phone began in the 1920s. The very first usage of it was in taxis/cars where it was used as a two-way radio for communication. Cell phones evolved over time like any other electronic equipment, and each stage or era was most certainly interesting. From its first official use by the Swedish police in 1946 to connecting a hand-held phone to the central telephone network, modern cell phones evolved tremendously. Ring (1947) created a communication architecture of hexagonal cells for cell phones. Later an engineer discovered that cell towers can both transmit and receive signal in three different directions led to further advancement. Early cell phone users were limited to certain blocks of area often referring to base stations covering a small land area. It was not possible to remain in reach beyond such defined boundaries until Joel's development of handoff system. By this, users were enabled to roam freely across cell areas without interruption to their calls. Cell phone had analog services between 1982 and 1990. In 1990, Advanced Mobile Phone Services (AMPS) turned the analog services to digital and went online ("History of Cell Phone" 2010).

1.1.1 First Generation (1G) Mobile Phones

The USA Federal Communication Commission (FCC) approved for public use the first cell phone called Motorola DynaTAC 8000X from Motorola but was made available to the public market after 15 years and was developed by Dr. Martin Cooper. It was considered to be a lightweight cell phone of about 28 ounces. Its dimensions were 13 x 1.75 x 3.5 inches. First generation mobile phones worked with the Frequency Division Multiple Access (FDMA) technology. The first generation mobiles are large in size and heavy to carry. First generation mobile phones were used only for the voice communication purpose("History of Cell Phone" 2010).

1.1.2 Second Generation (2G) Mobile Phones

The second generation mobile phones were introduced in the 1990s. Second generation (2G) mobile phones worked with both GSM and CDMA (Code Division Multiple Access) technologies.

2G network signals are digital while 1G network signals are analog. 2G cell phones were smaller, weighing between 100 to 200 grams; these were hand-held and were portable. Later improvements on these cell phones included faster internet access with GPRS (General Packet Radio Service) and subsequently, EGDE (Enhanced Data rates for Global Evolution) technology. And sharing of files with other mobile devices using infra red or Bluetooth technology. There were other improvements like Short Message Service (SMS) smaller batteries and longer battery life, etc. Due to all these improvements, the mobile phone customer base expanded rapidly worldwide.

1.1.3 Third Generation (3G) Mobile Phones

Most present day mobile phones are the third generation phones. The standards used on 3G phones differ from one model of the mobile phone to the other which essentially depends on the network providers. These phones were capable of streaming live videos, stream radio, making video calls, send emails, have mobile TV, have high internet access speed due to HSDPA (High Speed Data Packet Access) and WCDMA (Wideband Code Division Multiple Access) technology. They also use Wi-Fi and touch screen technology apart from performing all the functions of the 2G mobile phones ("History of Cell Phone" 2010).

1.2 Dual SIM Mobile Phones

A *dual-SIM mobile phon*e has the capacity to hold two SIM cards. The earliest model of this technology made use of dual-SIM adapters on *single SIM phones* which of course had only one transceiver. The use of this adapter rendered a slim phone bulky. Sometime the SIM card needed to be trimmed to fit into the adapter and the phone. The *dual-SIM adapter* could hold two SIMs at a time and was small enough to fit behind the battery of a regular mobile phone. However both SIMs could not be activated at the same time on the mobile phone. Switching from one SIM to the other was done by restarting the mobile phone; this combination is called a *standby dual-SIM phone*. Recent dual-SIM phones have both SIMs activated simultaneously such that there is no need to restart the phone, these are referred to as *active dual-SIM phones*. Most of these phones have two transceivers in built of which one of the transceivers may support both 2G and 3G while the other transceiver only supports 2G. Another type of dual-SIM mobile phone exists which supports both GSM and CDMA network. A new generation of dual SIM mobile phones makes use of only one transceiver yet provide 2 active SIMs simultaneously e.g. the LG GX200. Some dual-SIM phone use calls management software and can divert calls from one SIM to the other SIM's voicemail when a call is in progress or simply indicate that the line is busy. Both SIM share the phone's memory such that they share the same contact list and SMS and MMS message library (li, R. 2010). A recent introduction is mobile phones capable of holding three SIM cards; an example is the Akai Trio ("Dual SIM" 2011).

1.3 Telephony in Nigeria

Up until 2001 Nigeria experienced problems with the services provided by its then main communications service provider the Nigerian Telecommunication Plc (NITEL) including inefficient services, lack of access, limitation of services to places of institution since only landlines were majorly deployed.In 1992, the telecommunications industry in Nigeria was deregulated. First was the commercialization or corporatization of Nigerian Telecommunications Plc (NITEL) while the second was the establishment of the Nigerian Communications Commission (NCC), the telecommunications industry regulator (Alabi 1996). The deregulation led to the introduction of Global System for Mobile communication (GSM) network providers operating on the 900/1800

MHz Spectrum were MTN Nigeria, Econet (now Airtel), Globacom and Mtel in 2001. As a result the use of mobile phones soared, and has replaced the unreliable services of the Nigerian Telecommunications Limited (NITEL). With an estimated 45.5 million mobile phones in use as at August 2007, and most people having more than one cell phone, Nigeria has witnessed a phenomenal growth in this sector ("Telecommunications in Nigeria," 2011).

2.0 THE NEED FOR DUAL SIM MOBILE PHONE

The GSM service in Nigeria came with its own problems as subscribers were not getting value for their money. Tariffs were high and the GSM service providers were plagued with numerous problems such as instability in power supply, insecurity of infrastructure, call drops, difficulty in network accessibility. Due to the peculiar nature of power supply in Nigeria, GSM service providers had difficulty in powering their cell sites. Electric power generators installed at base stations to supplement or provide power meant additional deployment and operational cost. This in advertently led to increase in call tariffs.

GSM service providers also incurred additional cost with regard to securing installed facilities. GSM Service providers have high numbers of security personnel on their payroll, because these guards are needed to guard their installations against theft and vandals. As of October 2007, Airtel (formerly Zain) had 2500 base stations, MTN-2900, and Globacom-3000 in Nigeria (Adegoke, Babalola, et al 2008). With two security personnel per cell site, one can relate the cost. These costs go into the total cost of operation thereby leading to increases in call tariffs. The presence of security personnel doesn't however guarantee the safety of these facilities since there are reported cases of stolen generators and siphoned fuel from reservoirs (Njoku, 2007).

Major complaints from network subscribers were the inability to access the network to initiate calls. A subscriber had to dial several times before a call could go through. Sometimes after dialing several times, a subscriber might be connected to the wrong number.

Often established calls are abruptly terminated in the middle of conversions. This can happen for several reasons. There may be loss of signal between the network and the mobile phone, when the mobile phone (subscriber) is outside the network coverage area, or the call is dropped upon handoff between cells on the same provider's network. Other causes include cell sites running at full capacity no room for additional traffic, poor network configuration such that a cell is not aware of incoming traffic from a mobile device; the call is lost when the mobile phone cannot find an alternative cell to handoff.

2.1 The GSM Service

Network accessibility, dropped calls and high tariff appear to be must worrisome to the average GSM subscriber. A common maxim then was "of what use is a mobile phone when it cannot be used at will?" Disturbingly, GSM service network problems often persist for days and on rare occasions, for weeks. These problems are peculiar to all the service networks. When one network is down, often service may be available on other networks. The logical option to subscribers was subscription to multiple networks; this of course meant acquisition of multiple GSM phones, with the attendant inconvenience associated with having to keep more than one handset. Many subscribers using multiple handsets experienced loss or theft of some of these phones.

Most Nigerians therefore desired and looked for a means of having two SIMs on a phone to overcome the problem of carrying more than one phone. Major mobile phone manufacturers however concentrated on producing sophisticated phones with mind blowing features like camera, FM radio, memory card, WAP, GPRS and EGDE capabilities at the time. These companies excelled in *product performance* using current technology to produce better and more durable phones with each new release. They sustained *product performance* with their new innovations and high tech phones.

3.0 THE DISRUPTIVE IMPACT OF DUAL-SIM TECHNOLOGY

Mobile phones of all brands, shapes and sizes, were introduced into the phone market at the onset just as GSM service providers were expanding network coverage. Common household names included Nokia, Samsung, Sagem, Sony Erickson and LG. There were a few other brands albeit insignificant compared to these six. The trend was slick, high-tech mobile phones with improved performance and durability. However, Chinese phone manufacturing companies introduced a disruption in this market trend and became a major player on the *Nigerian mobile phone market* via the introduction of dual-SIM capable phones popularly called "China Phones". Although these products did not equal the existing brands in performance, look and durability, they provided an innovational intervention for the target market in providing access to multiple service networks on a single phone. As such, with the additional vantage of being cheap and easily affordable, the Nigerian market embraced the product. Most of the features on existing sophisticated phones are also available on the *dual-SIM phones*. According to a market research company GFK Retail and Technology, 30 per cent of mobile phones in Nigeria are dual-SIM (Rattue, 2011). This development which is directly related to the phenomenal growth of multi-SIM devices globally is not only in Nigeria. In Indonesia, Vietnam, Ghana and India, the market has grown from one in ten in 2009 to one in four by the quarter of 2010. According to the report, in Middle East and Africa, one in every 10 mobile phones sold uses dual SIM. In Asia, 16 per cent of all mobile phones sold have dual SIM capabilities, which represents an increase from 13 per cent at the beginning of 2010 (Rattue, 2011). There were however warranty issues with the first adapter type *dual-SIM phones*; these adapters could be used with normal *single SIM phones*. The use of the *dual-SIM adapte*r voided warranty for such phones. Also "*China Phones*" that are *active dual-SIM phones* are bought from the dealers without warranty. When asked "why?" they reply that they equally bought them wholesale without warranty. Another issue is that of durability; "*China Phones*" broke down unpredictably. In event of fault, local repair shops find it difficult to get replacement parts as there are no service centers or parts shop for such products. The lack of International Mobile Equipment Identity (IMEI) numbers in the unbranded made in China handsets makes them non-traceable and creates security concerns. In spite of these shortcomings the demand for them is ever increasing, as low income earners can easily afford them. Most local mobile phone outlets sell mostly these "*China Phone*". Established firm and global mobile phone manufacturers are facing stiff competition from Chinese brands and "fakes" in the Nigeria mobile phone market. This they have done by enticing consumers with attractive combination of features at affordable prices. Chief among these features is the dual-SIM capabilities of these mobile phones which established manufacturers are slowly introducing (Rattue, 2011). Samsung's D880 Duos was not so successful when it was introduced since calls were only possible with its primary SIM, unlike the Chinese brands which offered dual call capabilities. To initiate call from the secondary SIM of a Samsung D880, it must be made the primary. This difficulty in addition to its high cost were factors that made it unsuccessful. Subsequent Samsung *active dual-SIM phone* models had better performance but their costs were still high. Nokia only introduced their cheap dual-SIM phone, Nokia C Series in Nigeria in 2010. There is general acceptance in the country that Nokia phones are more durable than others. However, the Nokia C1-00 is a *standby dual-SIM phone* as only one of the SIM is active at a time.

Initially most Nigerians embraced the *dual SIM phones* due to inconvenience associated with carrying two mobile phones at the time. Presently, there is improvement in the power sector in the country. There have also been reductions in inaccessibility to GSM networks and the rate of dropped calls. And insecurity still remains an issue at each cell sites. The inclination to ownership of multiple mobile phones is currently not only driven by these factors but by new factors including lower call tariff, promo by various GSM service providers to entice customer and privacy/ personal security issues.

3.1 The Way Forward

Some of the problems facing the dual-SIM "China Phones" and possible steps to address them.

- No warranty Issues
- Poor durability
- No service centres
- Difficulty in Getting Replacement Parts
- Security Issues (No IMEI number)

The cases of void warranties as a result of using *dual-SIM adapter* on normal single SIM mobile has drastically reduced if not eliminated by recent *active dual-SIM mobile phones*. The lack of warranty for a product often creates doubt in the mind of the customers as to durability or authenticity of the product. Wholesale dealers who get these "*China Phones*" should be made to demand that the manufacturers of such phones issue warranties for them. This will encourage more patronage.

The issue of durability can be a result of poor design or the use of substandard materials to implement the *dual-SIM technology*. Since most of these phones are cheap as compared with other *dual-SIM mobile phones* manufactured by big and popular mobile phone manufacturers like Nokia, Samsung etc. There is the likelihood that the use of substandard materials is the cause of poor durability. Better materials will lead to increase in cost of production and product cost. I believe they can strike a balance and still produce phones that are reasonably priced. Initially, the "*China Phones*" had short battery life, but the phones now come with extra battery.

Chinese phone manufacturing companies need to establish service centres in the country or train and certify a hand full of owners of local mobile phone repair shop who will in turn pass on the skills acquired to others such that there will be enough skilled technicians who will be able to repair these phones in the event of faults. Replacement parts for "*China Phones*" should be made available to the trained technicians through the service centres.

There is need for regulation to stop the use of *dual-SIM mobile phones* without IMEI number. International Mobile Equipment Identity (IMEI) number is unique to every GSM and WCDMA mobile phone and found printed inside the battery compartment of the phone. It can be displayed on the screen of the phone by entering*#06# on the keypad.

In India when a large percentage of people used such phones, mobile operators implanted IMEIs onto such phones rather than bar services. But the Indian government placed a ban on the usage of phones without IMEI which took effect from December 1, 2009.

4.0 CONCLUSION

The need for communication in spite of poor network coverage and quality of service by the Mobile (GSM) service providers, informed ownership of multiple number of *single SIM mobile phones* to guarantee access to available network services by mobile phone users in Nigeria, with associated multi-phone ownership problems. Major mobile phone manufacturers who prefer a *sustained technology* model responded to the increase of market with improved and more sophisticated products. Disruptive market innovation by way of *dual-SIM mobile phones* products met market anticipation. These dual-SIM " made in China" phones were not as attractive or as durable as the existing sophisticated brands but had most features on these phones and were cheap and affordable. The dual-SIM phones performance were also affected by short battery life, plus absence of warranties, technical support/services outlets and replacement parts. Additional security issues are associated with the phones' lack of IMEI numbers. However, the *dual-SIM innovation* met a market need and is widely used in Nigeria. Whereas electricity supply, insecurity and other problems which informed telecoms service quality which informed multiple phone ownership are declining, personal security issues and preference of lower tariff offerings continue to inform multiple network access. As such, dual SIM phones remain a popular market choice. Associated problems with this dual SIM products must however be addressed by the China based manufacturers and other market players.

REFERENCES

Adegoke, A. S., Babalola, I. T., & Balogun, W. A. (2008). Performance evaluation of GSM mobile system in Nigeria. *Pacific Journal of Science and Technology*, *9*(2), 436–441.

Alabi, G. A. (1996). Telecommunications in Nigeria. Retrieved March 10, 2011, from www.africa.upenn.edu

Christensen, M. C. (1997). *Innovator's dilemma*. Harvard Business School Press.

Christensen, M. C., & Raynor, M. E. (2003). *Innovator's solution*. Harvard Business School Press.

Dual, S. I. M. (2011). Retrieved March 10, 2011, from http://en.wikipedia.org/wiki/Dual_SIM

Dual SIM Mobile Phones. (2009). Retrieved March 10, 2011, from http://www.dualsimmobilephones.com/2009/09/dual-sim-mobile-phones/

History of Cell Phone. (2010). Retrieved March 10, 2011, from www.historyofcellphones.net/

Li, R. (2010). *Cell phone mysteries, what is dual-SIM?* Retrieved March 9, 2011, from www.articles.webraydian.com

Njoku, C. (2007). *The real problem with GSM in Nigeria*. Retrieved March 9, 2011, from http://www.nigeriavillagesquare.com/index2.php?option=com_content&do_pdf=1&id=7829

Rattue, A. (2009). Buoyant Nigerian market sees 15 million mobile handsets sold in 2009. Retrieved July 12, 2011, from http://www.gfkrt.com/news_events/market_news/single_sites/005203/index.en.html

Rattue, A. (2011). Multi SIM phenomenon continues in emerging mobile markets. Retrieved July 12, 2011, from http://www.gfkrt.com/news_events/market_news/single_sites/007260/index.en.html

Sally, M., Sebire, G., & Riddington, E. (2010). *GSM/EDGE: Evolution and performance* (p. 504). John Wiley and Sons Ltd. doi:10.1002/9780470669624

Telecommunications in Nigeria. (2011). Retrieved March 10, 2011, from http://en.wikipedia.org/wiki/Telecommunications_in_Nigeria#mw-head

KEY TERMS AND DEFINITIONS

Active Dual-SIM Phone: A dual-SIM that has both SIM activated, calls can be made or received simultaneously and there is no need to restart or switch between SIMs.

Cell Phone: America's name for mobile phone

China Phone: substandard and sometimes unbranded dual-SIM mobile phone manufactured in China

Dual-SIM Phone: Mobile phone capable of holding two SIM cards that may or may not have both SIM cards activated to make or receive calls simultaneously.

Standby Dual-SIM Phone: A dual-SIM mobile that has one SIM activated at a time and needs to be restarted to activate the other SIM or switch between SIMs.

This work was previously published in Disruptive Technologies, Innovation and Global Redesign: Emerging Implications, edited by Ndubuisi Ekekwe and Nazrul Islam, pp. 462-469, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 36
Data Envelopment Analysis in Environmental Technologies

Peep Miidla
University of Tartu, Estonia

ABSTRACT

Contemporary scientific and economic developments in environmental technology suggest that it is of great importance to introduce new approaches that would enable the comparison of different scenarios for their effectiveness, their distributive effects, their enforceability, their costs and many other dimensions. Data Envelopment Analysis (DEA) is one of such methods. The DEA is receiving an increasing importance as a tool for evaluating and improving the performance of manufacturing and service operations. It has been extensively applied in performance evaluation and benchmarking of several types of organizations, so-called Decision Making Units (DMU) with similar goals and objectives. Among these are schools, hospitals, libraries, bank branches, production plants, but also climate policies, urban public transport systems, renewable energy plants, pollutant emission systems, environmental economics, etc. The purpose of this chapter is to present the basic principles of the DEA and give an overview of its application possibilities for the problems of environmental technology.

INTRODUCTION

The Earth is the common home for all of us and because of this the great attention paid to environmental problems is more than natural and urgent. The lack of economic value of environmental goods often leads to over-exploitation and degradation of these resources. It is extremely important to monitor and control interactions between production technologies and the environment. To keep and conserve the natural environment, environmental technology is developed. Independently of application areas of environment sciences, new approaches and methods, particularly of mathematical modeling, are extremely needed and welcome in this area. It is well-known that mathematical modelling is the most efficient method for investigating different processes, their simulation and prediction.

DOI: 10.4018/978-1-4666-1945-6.ch036

Data Envelopment Analysis (DEA) is a relatively new data oriented mathematical method for evaluating the performance of a set of peer entities traditionally called Decision Making Units (DMU) which convert multiple inputs into multiple outputs. Since DEA was first introduced, in its present form, in 1978 (Charnes et al., 1978), we have seen a great variety of applications of DEA being used in evaluating the performances of many different kinds of entities engaged in many different activities in many different contexts in many different countries. These DEA applications have used DMUs of various forms, such as hospitals, schools, universities, cities, courts, business firms, and others, including the performance of countries, regions, etc. The DEA is frequently applied in many areas of applied economic sciences, including agricultural economics, development economics, financial economics, public economics, macroeconomic policy, etc.; and among others, in addition to its traditional confinements in productivity and efficiency analysis; it has also diffused into the field of environmental economics and environmental technology. As it requires very few theoretical assumptions, DEA has opened up possibilities for its use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in DMU.

There are several examples of areas of environmental science, where DEA is used and where remarkable theoretical and practical results have been achieved. In the framework of problems raised by climate change, one of the major threats to the Earth's sustainability, we see that DEA is applied for assessing the relative performance of different climate policy scenarios as DMUs, through accounting their long-term economic, social and environmental impacts as input and output parameters. (Bosetti & Buchner, 2008) Quantitative techniques are crucial here in order to make adequate decisions quickly. In the sense of climate change measuring the carbon emission performance is also important. (Zhou et al., 2008) In another context, the paper by (Piot-Lepetit, 1997) considers the usefulness of DEA for estimating potential input reductions and assessing potential reductions of environmental impact on agricultural inputs. We can see the same in the paper (Madlener et al., 2006), where the assessment of the performance of biogas plants is realized. In the study (Rodrigues Diaz et al., 2004) DEA was used to select the most representative irrigation districts in Andalusia. One can find the use of DEA to assess corporate enactment of Environmentally Benign Manufacturing as work parts move from place to place in a company (Wu et al., 2006); this work touches green manufacturing problems. The DEA is used even in political decision making (Taniguchi et al., 2000), and to discuss a methodology to assess the performances of tourism management of local governments when economic and environmental aspects are considered as equally relevant. (Bosetti et al., 2004)

The structure of the present chapter is the following. First an overview of DEA method is given. Today the DEA itself has developed and has several forms, versions and modifications, each of which has specific application features. Below we formulate only the basic version of it because there is a lot of literature available in libraries and on the internet. The main part of the paper deals with the case studies, the results of which have been achieved by using the DEA. There is a rising trend to apply DEA and naturally some selection of them is included. Finally the reader finds the conclusions and references. One important objective of this chapter is to emphasize that environmental technologies are very open to innovation, and using new methods of mathematical modeling is a part of this.

OVERVIEW OF THE DATA ENVELOPMENT ANALYSIS

In this paragraph we give a short description of Data Envelopment Analysis (DEA). More profound treatments of the topic can be found in many books, (e.g. Cooper et al., 2004; Cooper et al., 2007; Thanassoulis 2003) and there are also books (edited by Charnes et al., 1994) and papers dealing with the applications of DEA. Comprehensive information about DEA can be found on the web-page of Ali Emrouznejad (Emrouznejad 1995-2001) or in the paper (Emrouznejad et al., 2008). Also, it is important to mention that all the papers referred to in the next section "Case Studies" contain a sufficient overview of DEA versions or modifications used in particular cases under consideration.

DEA and Benchmarking

Data Envelopment Analysis belongs to the wider class of efficiency measuring methods, to the so-called frontier methods, and is a data oriented approach for evaluating the performance of a set of peer entities called Decision Making Units (DMUs). DEA is a multi-factor productivity analysis model for measuring the relative efficiencies of DMUs and is receiving increasing importance as a tool for evaluating and improving the performance of manufacturing and service operations. DEA is a powerful methodology for identifying best practice frontier and has been extensively applied in performance evaluation and benchmarking of schools, hospitals, bank branches, production plants, public-sector agencies, etc. Technically DEA represents the collection of non-parametric, linear programming techniques developed to estimate the relative efficiency of DMUs. Largely the result of multi-disciplinary research in economics, engineering and management, DEA is a theoretically sound framework that offers many advantages over traditional efficiency measures such as performance ratios and regression.

The most important feature of DEA is its ability to handle effectively the multidimensional nature of inputs and outputs in the production and management process. Efficiency as an economic category is in use widely in profit – targeted organizations and enterprises. The simple idea that greater input into system brings us greater output and increases the profit lies in the base of the general meaning of efficiency. Theoretically and practically several types of efficiency are in use and every one of them gives a different possibility for interpretation and for making new decisions. The same is also seen in DEA concepts. Here for each DMU the efficiency score in the presence of multiple input and output factors is defined as: weighted sum of outputs divided to weighted sum of inputs. All scores are between zero and one; DMU's whose efficiency is equal to one, are certified as fully efficient or simply efficient. For every inefficient DMU, i.e. for these which efficiency score is less then one, the DEA identifies a set of corresponding efficient units, so called reference group, that can be utilized as benchmarks for the improvement of activities. The estimation of the efficiency of a single DMU, on the base of other organizations acting in the same economic environment, is the main advantage of the DEA approach. Reference group means: the set of real DMUs whose outputs and inputs belong to the outputs and inputs of composite virtual DMU with nonzero weights. The fact that efficiencies of reference DMUs are taken equal to 1 (or 100%) does not mean that the quality of work of those base organizations could not improved. The results of the application of the DEA fix only the situation in the moment when the data for input and output parameters is collected.

Geometrically for every DMU evaluated the reference set determines a frontier in the input – output parameters space. This is called best practice production frontier and the points which correspond to efficient DMUs are situated on this. If in the group of DMUs under consideration are **n** members for all the values of demanded input-output parameters, then in the parameters space we have also **n** data points. The best practice frontier is constructed then as piecewise linear surface in this space which envelopes set of data points and separates it from the beginning of coordinates. Fully efficient DMUs are vertices of this. Efficiency means geometrically a radial measure which quantitatively equals to the rate of distances from the beginning of coordinates of the composite virtual DMU and DMU being evaluated. Evaluation process of every single nonefficient DMU gives in general different efficiency coefficients and individual reference set because the data points are different. Exceptions are efficient DMUs: as mentioned, the efficiency of a DMU which belongs to reference set is established equal to one but over that there may be other DMUs which efficiency is also equal to one but which do not belong to any reference set. In all textbooks one can find the exact explanations also about geometrical issues. (See e.g. Cooper, et al., 2004).

Data for DEA is a set of parameters, evaluated for each DMU. The parameters are divided into input and output parameters which represent different activities of DMUs under consideration. For example, if we consider DMUs as public libraries, the input parameters could be fixed as yearly expenditures on acquisition of new books, yearly expenditures on salaries, the size of collections and the area of the library rooms. As for output parameters one may consider the number of readers and the number of loans. It is interesting to mention some important conclusions of this work although research considered libraries. In the years 2002 by 2005 eight central libraries, i.e. 40% from the whole selection, used their resources effectively. The score of relative efficiency of the rest of libraries varied from 0,740 to 0,979.The data of four years show that the trend of the score of efficiency of central public libraries of Estonia was falling, i.e. the average score decreases. In 2005 there were six central libraries from 20 which were scale efficient, i.e. of optimal size. (Miidla & Kikas, 2009). The list of parameters and division into inputs and outputs can be chosen and fixed differently, depending on the particular goals of research and, naturally, on application area. One can say that DMUs convert multiple inputs into multiple outputs. It is just interesting that the inputs and outputs do not need to have a common measure; they can be quantities of completely different units – meters, dollars, number of persons, number of lakes in regions, etc. It is important that in the whole selection the inputs and outputs of all DMUs have a value; although in the case of absent data there are also approaches which allow the use of DEA (Kao & S-Tai Liu, 2000). Further with regards to the words ‘selection’ or ‘selection group’ we refer to the whole set of DMUs under consideration.

The DEA approach uses a linear programming model for measuring the relative efficiencies of those DMUs on the basis of the given data. First, a DMU is fixed and a hypothetical composite operating unit, a hypothetical DMU, based on all units in the selection group, is constructed. The input and output of this composite virtual DMU are determined by computing the weighted average of the inputs and outputs of all real DMUs in the selection group, and the efficiency score of the initially fixed DMU is defined as the rate output and input of this constructed composite DMU. This procedure is repeated for each single DMU, and as the output for the DEA application itself one gets an array of these relative efficiencies, which lay between zero and one. Thus the DEA approach is a kind of peer comparison method. Constraints in the linear programming model require all outputs of the composite virtual DMU to be greater than or equal to the outputs of the DMU being evaluated. So, if in the selection group there are **n** members then for evaluating all of the members, from the point of view of relative efficiency, we need to establish and solve **n** problems of linear programming. It should be mentioned that often more than one problem of linear programming is needed to solve each DMU. In the advance use of DEA, it is necessary to use more than one problem of linear programming when a more detailed analysis is needed. (Thanassoulis, 2003)

If the inputs for the virtual composite unit can be shown to be less than the inputs for the DMU being evaluated, the composite DMU will be shown to have the same, or more output for less input. In this case, the model will show that the composite virtual DMU is more efficient than the DMU being evaluated. In other words, the DMU under evaluation is less efficient than the virtual DMU. Since the composite DMU is based on all DMUs in the selection group, the DMU being evaluated can be judged relatively inefficiently when compared to the other units in the selection. The estimation of the efficiency of a single DMU on the basis of others acting in the same environment, the so-called reference group, is the main advantage of the DEA approach. The reference group means the set of real DMUs whose outputs and inputs belong to the outputs and inputs of composite virtual DMU with nonzero weights. This approach makes the DEA very attractive because in the case of environmental entities it is difficult to speak about efficiency for all these in one sense, as it is possible, for example, for profit organizations. The efficiency of reference DMUs is taken equal to 1 (or 100%). The results of the application of the DEA can only fix the situation when applying the results to the time frame in which the data for input and output parameters was collected. The environment may have already changed; applying these results to the following year can render inaccurate data.

Geometrically for every DMU evaluated the reference set determines the frontier in the input – output parameters space. This is called the best practice production frontier, although in the case of environmental technologies it is difficult to speak so, and the points which correspond to effective DMUs are situated on this. As mentioned before, the frontier itself is a piecewise linear surface in the input - output parameters space as an envelope to a production possibility set. Points corresponding to the reference or fully efficient DMUs are vertices of this frontier. Efficiency is geometrically a radial measure quantitatively equal to the rate of distances from the beginning of coordinates of the composite virtual DMU and of DMU being evaluated. This rate is just referred to as the relative efficiency. Evaluation of every single DMU gives in principle a different frontier and an individual reference set. The efficiency of DMUs in the reference set is established equal to one and there may be also other DMUs which efficiency is equal to 1 which do not belong to any reference set, they belong to the composite virtual DMU with zero weight. The best practice frontier relies only on the initial data, i.e. on inputs and outputs of all DMUs in the selection. The algorithm of efficiency estimation does not require the construction of the best practice frontier; the numerical method gives the answer without geometrical interpretation.

Considering the DEA methodology one has to assume scaling properties of input parameters, returns to scale. This means the influence of inputs variation to outputs changes. Let us assume that the input parameters are changed, all in the same proportionality. If the outputs change in the same proportion we can speak about constant returns to scale. Otherwise, when outputs do not change in the same rate we can speak about variable returns to scale; more precisely, increasing or decreasing returns to scale if outputs change in a greater or smaller proportion. The scale issue associated with DEA might be under attention when different DEA models are in use and the scaling property leads us to the meaning of optimal size of DMU. Namely, the DMU is shown to be of optimal size when it is efficient both in the sense of constant returns to scale and variable returns to scale. If the DMU is less than optimal, it usually works in the conditions of increasing returns to scale. Conversely, oversized, if compared to the optimal size DMUs, works in the conditions of decreasing returns to scale. This might also be important in some aspects of application in environmental technology.

The DEA approach can be input-oriented or output-oriented. In the first case the main question is: how much is it possible to decrease the input parameters for inefficient DMUs in order to keep the present output. In the case of output-oriented method the main question is: how much is it possible to increase the outputs while keeping the present input. Choosing between these two possibilities depends again on the context of application, both of the cases lead to different modifications of the DEA method.

Mathematical Formulation

In their initial study, Charnes, Cooper, and Rhodes (Charnes et al., 1978) described DEA as a 'mathematical programming model applied to observational data that provides a new way of obtaining empirical estimates of relations, such as the production functions and/or efficient production possibility surfaces, that are cornerstones of modern economics.' In this article they proposed following the fractional program model, known as the CCR DEA model.

Assume that there are **n** DMUs to be evaluated. Each DMU consumes varying amounts of **k** different inputs to produce **m** different outputs. Specifically, DMU **j** consumes amount $\mathbf{x_{ji}}$ of input **i** and produces amount $\mathbf{y_{jr}}$ of output **r**. We assume that $\mathbf{x_{ji} \geq 0}$ and $\mathbf{y_{jr} \geq 0}$ and also assume without loss of generality that each DMU has at least one positive input and one positive output value. The first approach gives us the fractional program problem for evaluating the efficiency score of the DMU number **s**. In this form, as introduced by Charnes, Cooper, and Rhodes, the ratio of outputs to inputs is used to measure the relative efficiency of the DMU to be evaluated relative to the ratios of all of the **n** DMUs. We can interpret the CCR construction as the reduction of the multiple-output/multiple-input situation for each DMU to that of a single virtual output and virtual input. For a particular DMU the ratio of this single virtual output to single virtual input provides a measure of efficiency that is a function of the multipliers $\mathbf{v_1},\ldots,\mathbf{v_m}$, $\mathbf{u_1},\ldots,\mathbf{u_k}$. In mathematical programming parlance, this ratio, which is to be maximized, forms the objective function for the particular DMU being evaluated. Symbolically the problem is the following:

Find

$$\mathbf{max}\ \{(\mathbf{v_1\, y_{1s}} + \ldots + \mathbf{v_m\, y_{ms}})/(\mathbf{u_1\, x_{1s}} + \ldots + \mathbf{u_k\, x_{ks}})\},$$

maximization over

$$\mathbf{v_1},\ldots,\mathbf{v_m},\ \mathbf{u_1},\ldots,\ \mathbf{u_k}$$

subject to:

$$\{(\mathbf{v_1\, y_{1i}} + \ldots + \mathbf{v_m\, y_{mi}})/(\mathbf{u_1\, x_{1i}} + \ldots + \mathbf{u_k\, x_{ki}})\} \leq 1,\quad \mathbf{i} = 1,\ldots,\mathbf{n}$$

$$\mathbf{v_1},\ldots,\mathbf{v_m},\ \mathbf{u_1},\ldots,\ \mathbf{u_k} \geq 0$$

where $\mathbf{v_1},\ldots,\mathbf{v_m}$, $\mathbf{u_1},\ldots,\mathbf{u_k}$ are weights given to real observed output values $\mathbf{y_{1s}}, \ldots, \mathbf{y_{ms}}$ and input values $\mathbf{x_{1i}}, \ldots, \mathbf{x_{ki}}$, correspondingly.

In this first approach, the fractional program problem has an infinite number of solutions:

if

$$(\mathbf{v_1}^*,\ldots,\mathbf{v_m}^*,\ \mathbf{u_1}^*,\ldots,\ \mathbf{u_k}^*)$$

is a solution of this, i.e. optimal, then also

$$(\boldsymbol{\alpha}\,\mathbf{v_1}^*,\ldots,\ \boldsymbol{\alpha}\,\mathbf{v_m}^*,\ \boldsymbol{\alpha}\,\mathbf{u_1}^*,\ldots,\ \boldsymbol{\alpha}\,\mathbf{u_k}^*)$$

has the same property for every $\boldsymbol{\alpha} > 0$. The additional condition

$$\mathbf{u_1\, x_{1s}} + \ldots + \mathbf{u_k\, x_{ks}} = 1$$

makes it possible to convert this fractional problem to following linear programming problem for the estimation of the efficiency of DMU number **s**.

Find

$$\mathbf{z^*} = \max (\boldsymbol{\mu}_1 \mathbf{y}_{1s} + \ldots + \boldsymbol{\mu}_m \mathbf{y}_{ms}),$$

maximization over

$$\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m$$

subject to:

$$(\boldsymbol{\mu}_1 \mathbf{y}_{1i} + \ldots + \boldsymbol{\mu}_m \mathbf{y}_{mi}) - (\boldsymbol{\gamma}_1 \mathbf{x}_{1i} + \ldots + \boldsymbol{\gamma}_k \mathbf{x}_{ki}) \le 0, \quad \mathbf{i} = 1,\ldots,\mathbf{n}$$

$$\boldsymbol{\gamma}_1 \mathbf{x}_{1s} + \ldots + \boldsymbol{\gamma}_k \mathbf{x}_{ks} = 1$$

$$\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k \ge 0.$$

This is a basic linear programming problem for obtaining the efficiency score **z*** of the DMU number **s.** The above problem needs to run **n** times in identifying the relative efficiency scores of all the DMUs. Each DMU selects input and output weights that maximize its efficiency score. As mentioned, a DMU is considered to be efficient if it obtains a score of 1; a score of less than 1 implies that it is inefficient. When working with literature, a reader can find several versions of DEA computational schemes and should clearly understand what version is in use in every particular case, particularly understand the assumptions, etc. In this paper one does not find any exhaustive discussion but simply an example of DEA method formulation. Below we bring the dual problem of linear program.

Find

$$\boldsymbol{\Theta^*} = \mathbf{min}\ \boldsymbol{\Theta},$$

minimization over weights

$$\lambda_1, \ldots, \lambda_n$$

subject to:

$$\boldsymbol{\lambda}_1 + \ldots + \boldsymbol{\lambda}_n = 1$$

$$\boldsymbol{\lambda}_1 \mathbf{x}_{1,i} + \ldots + \boldsymbol{\lambda}_n \mathbf{x}_{n,i} \le \boldsymbol{\Theta} \mathbf{x}_{s,i}\ , \ \mathbf{i} = 1,\ldots,\mathbf{k}$$

$$\boldsymbol{\lambda}_1 \mathbf{y}_{1,j} + \ldots + \boldsymbol{\lambda}_n \mathbf{y}_{n,j} \ge \mathbf{y}_{s,j}\ , \ \mathbf{j} = 1,\ldots, \mathbf{m}$$

$$\boldsymbol{\lambda}_1 \ge 0,\ldots, \boldsymbol{\lambda}_n \ge 0.$$

As in the case of the main problem, for a complete realization of the DEA it is necessary to solve the linear programming problem for every single DMU in the selection. This gives us the array of relative efficiencies, solutions **Θ*** for every single DMU and only then it is possible to start with the interpretation process. Given dual problem is important because the units involved in the construction of the composite DMU, i.e. for which corresponding weight $\boldsymbol{\lambda}_i > \mathbf{0}$, can be used as benchmarks for improving the inefficiency of tested DMU's. DEA allows for computing the necessary improvements required in the inefficient DMU's inputs and outputs in order to make it efficient. If a DMU is inefficient, the corresponding solution **Θ*** is less than one; the solution also gives us the parameters of the corresponding composite virtual DMU which is efficient, of course, i.e. the weights for its input and output. This virtual DMU also presents us the corresponding point on the production efficiency frontier; 'dragging' the inefficient DMU point to the frontier means a decrease in the inputs of this DMU.

Sometimes it so happens that the composite DMU is located on the part of the production efficiency frontier which is parallel to some coordinate plane. In this case it is possible to realize an additional shift of the virtual data point of composite DMU down along of this parallel part towards the beginning of coordinates without decreasing outputs and increasing other inputs. This surplus is called a slack and it gives additional information about the inefficiency of the DMU under consideration. Slacks show the real possibilities to decrease the corresponding input parameter on the existing real example of other DMU working in such conditions. The interpretation of inefficiency in the case of input-oriented DEA, represented by dual linear programming problem, is easy. Namely, if we have **k** input parameters under consideration and if the relative efficiency Θ^* of a DMU is less than one, $\Theta^* < 1$, then to reach the full relative efficiency $(\Theta^* = 1)$ we must decrease inputs $(x_1,\ldots,x_k)$ of this DMU Θ^* times, i.e. to the values $(\Theta^* x_1,\ldots, \Theta^* x_k)$.

Any further discussion of these questions does not fall under the goal of the present paper. It is interesting to mention that in numerous applications of environmental technology, the versions and possibilities of Data Envelopment Analysis have been used in a great variety. The trend is increasing; the full DEA bibliography contains more than 4000 journal articles (Emrouznejad, 1995-2001) comprising many scientific fields.

CASE STUDIES

Next we explore the possibilities of DEA application in different areas of environmental technology. The use of DEA as a quantitative non-parametric performance measurement technique, based on linear programming, for assessing the relative efficiency of homogeneous peer entities called Decision Making Units (DMUs), is successfully implemented into environmental research projects. DEA makes it possible to compare the efficiency of each DMU to that of an ideal operating unit, rather than to the average performance. This ideal unit is constructed only on the base of the data for the whole set of DMUs and because of this several DMUs always become fully efficient. The definition of a DMU is generic and flexible and this enforces the dissemination of DEA into new areas. We explain the use of DEA on the base of some examples and most certainly not all application areas are covered. So, the overview given in this paper is in no way exhaustive of the developments in the field of DEA possibilities. The goal is to show some approach details, possible choices of DMUs, the corresponding input and output parameters, and the conclusions made on the basis of DEA. The interpretations of environmental efficiency in different cases are of interest. The order of examples is of no importance, as it is impossible to line up urgent environmental protection problems.

Climate Policy Scenarios

DEA is extended from its traditional application in (Bosetti & Buchner, 2008) as a quantitative method to assess the relative performance of different climate policy scenarios when accounting for their long-term economic, social and environmental impacts. Indeed, contemporary developments in the political, scientific and economic debate on climate change suggest that it is of critical importance to develop new approaches, particularly quantitative, to compare policy scenarios for their environmental effectiveness, their distributive effects, their enforceability, their costs and many other dimensions. As for input parameters for DEA application economic, environmental and social costs for every possible policy, also indicators of instant climate situation are considered here. Instead, among the outputs, there are indicators for which lower values are preferred and indicators of benefits and welfare. The authors discuss eleven simulated climate policy scenarios, computed three indicators for each of them (cumulated discounted GDP over the century, temperature increase by 2100, and the Gini equity indicator by 2100), add the application of DEA and make interesting conclusions. Two alternative DEA approaches to compare the sustainability of different policy scenarios are used. One of them is based on the efficiency score defined as a relative ratio and the other is based on Competitive Advantage measured in terms of absolute prices. The first case fits the traditional DEA application described above. Relative efficiency estimates are computed for each policy, where efficiency is measured as the ratio of the weighted sum of outputs to the weighted sum of inputs, and are obtained through solving a series of linear programming problems. Both, constant returns to scale and variable returns to scale assumptions are used. The interpretation of relative efficiency, computed using the DEA method, is interesting: the policy is 100% efficient if and only if 1) none of its outputs can be increased without either increasing one or more of its inputs, or decreasing some of its other outputs; 2) none of its inputs can be decreased without either decreasing some of its outputs or increasing some of its other inputs. In the second approach DEA is applied in order to get weights while for each scenario the net economic impact, expressed in monetary value, is aggregated through weights to the social and environmental impacts as DMUs and their efficiencies are expressed in their own unitary measures, on the base of the data from real activity.

Three major findings are pointed out: 1) stringent climate policies can outperform less ambitious proposals if all sustainability dimensions are taken into account; 2) a carefully chosen burden-sharing rule is able to bring together climate stabilization and equity considerations; 3) the most inefficient strategy results from the failure to negotiate a global post-2012 climate agreement. In conclusions the simulated scenarios, also the interpretational role of the DEA, are discussed in details. It is remarkable that it is possible to support the political, scientific and economic debate on climate change using the DEA method.

Measuring Environmental Performance Index

The Environmental Performance Index (EPI) is a method of quantifying and numerically benchmarking the environmental performance of a country's policies and has been in recent years universally adopted and quoted by corresponding policy analysts and decision makers. The construction of an aggregated EPI, which offers condensed information on environmental performance, has evolved as an important focus in systems analysis. Among the existing approaches to developing EPIs, some are data-driven while others are theory-driven. In the article of Zhou et al., (2008) we see an example of direct approach, where an aggregated EPI is directly obtained from the observed quantities of the inputs and outputs of the environmental system studied using Data Envelopment Analysis.

This work is an example of application of environmental DEA technology, in which outputs are assumed to be weakly disposable (Färe et al., 2004) and which have been widely used to measure industrial productivity when undesirable outputs exist. In recent years this approach has gained popularity in environmental performance measurement due to its empirical applicability. The common procedures for applying DEA to measure environmental performance are to first incorporate undesirable outputs in the traditional DEA framework, and then calculate the undesirable outputs orientation (environmental) efficiencies. In fact, many studies have been devoted to modeling undesirable factors in DEA, e.g. the data translation approach (Seiford & Zhu, 2005) and the utilization of environmental DEA technology.

In the article (Zhou et al., 2008) different DEA methods for environmental performance measurement are described (constant returns to scale and variable returns to scale, non-increasing returns to scale) and a study on measuring the carbon emission performance of eight world regions is presented. The centre of attention is on the growing concern on global climate change due to carbon dioxide (CO_2) emissions worldwide. The single input, desirable output and undesirable output are: total energy consumption (Mtoe, megatonne of oil equivalent), GDP (gross domestic product, billion 1995 US$) and CO_2 emissions (Mt), respectively.

Eight regions under consideration (i.e. DMUs for DEA application) are: OECD, Middle East, Former USSR, Non-OECD Europe, China, Asia (excludes China), Latin America and Africa.

In this study all the proposed models are radial DEA-based models where efficiency scores are computed as rate of distances from origin of coordinates, DEA frontier and data point of corresponding DMU. However, in some circumstances it may be difficult to compare some DMUs only by the proposed environmental performance indexes because of the weaker discriminating power of radial DEA efficiency measures compared to with non-radial DEA models. Authors propose to incorporate different environmental DEA technologies with the non-radial DEA efficiency scores and to combine different environmental DEA technologies with slacks-based efficiency measures. The results of the study are interesting. For instance, if the pure EPI is chosen and the reference technology exhibits variable returns to scale, OECD has a better carbon emission performance than Africa eventhough it has a larger carbon intensity and carbon factor.

Greenhouse Gas Technologies

In the paper (Lee et al., 2008), the authors use the Analytic Hierarchy Process (AHP) and Data Envelopment Analysis (DEA) hybrid model to weigh the relative preferences of greenhouse gas technologies in Korea. The Analytic Hierarchy Process is a subjective method used to analyze qualitative criteria in order to generate a weighting of the operating units and is known as a decision-making method which could be used to solve unstructured problems. In general, decision making involves tasks such as planning, the generation of a set of alternatives, the setting of priorities, the selection of the best policy once a set of alternatives has been established, allocation of resources, determination of requirements, prediction of outcomes, designing of systems, measurement of performance, ensuring of system stability, and the optimization and resolution of conflicts. Authors of the paper employed a long-term perspective when establishing the criteria to evaluate energy technology priorities for the greenhouse gas plan. They used the AHP to generate the relative weights of the criteria and alternatives in the greenhouse gas plan. Thereafter, the relative weights were applied to the data used to measure the efficiency of the DEA method. This study represents an example in which the AHP/DEA hybrid model has been used to determine the energy technology priorities for the greenhouse gas plan. The results obtained using this hybrid model provide the government with an effective decision-making tool and also represent a consensus of experts in the greenhouse gas planning sector.

Nine greenhouse gas technologies were considered as DMUs for DEA application: CO_2 capture storage and conversion technology, non-CO_2 gas technology, advanced combustion technology, next-generation clean coal technology, clean petroleum and conversion technology, DME (di-methyl ether) technology, GTL (gas to liquid) technology, gas hydrate technology and greenhouse gas mitigation policy. The parameters of DEA consist of a single input factor and multiple output factors. The input factor consists of the investment cost associated with the development of greenhouse gas technologies; the unit of investment cost was million US dollars in 2006. There are five output factors, namely possibility of developing technology, potential quantity of energy savings, market size, investment benefit, and ease of technology spread. All outputs are multiplied by the relative weights calculated using the AHP approach (these concern United Nations Framework Convention on Climate Change, UNFCCC, economic spin-off, technical spin-off, urgency of technology development, and quantity of energy use) and are thus applied in conjunction with the output factors employed as part of the DEA approach. As a result of the application of the AHP/DEA approach, one greenhouse gas technology, namely non-CO_2 gas technology with efficiency score 1, was found to be more efficient than the other eight greenhouse gas technologies. In conclusion, this hybrid model can be used to efficiently compute the relative efficiency scores of greenhouse gas technologies. This paper also shows decision makers and policy makers in the energy sector that multi-criteria decision making problems can be solved using scientific procedures.

Tourism Management

DEA is also effectively used for assessing the performances of tourism management of local governments when economic and environmental aspects are considered as equally relevant. In (Bosetti et al., 2004) the focus is on the comparison of Italian municipalities located in coastal areas. In order to assess the efficiency status of the considered management units, DEA is applied. In this analysis, the DMU represents a municipality producing the tourism good for which two different inputs are given. The first is the cost of managing the tourism infrastructure. More exactly, the cost of the production of tourism services is fixed on the total number of beds, which is considered as an approximation for management expenses, and it is computed by adding up the number of all beds in hotels, camping, registered holiday houses and other receptive structures. The second input is the environmental cost deriving from the increased number of people depending on the same environmental endowment; this parameter is presented as the amount of solid waste in tons per year. As output parameter, an indicator measuring the rate of use of existing beds has been used as a general approximation for profit deriving from the tourist industry, is considered.

The authors confirm that in the present study output-oriented models have been preferred to input-oriented ones, as they suit more issues considered as relevant for management purposes and they better help in addressing the germane questions. These applications also give the sense to the input and output indicators because in order to augment the efficiency of an inefficient municipality, the most direct policy means is to introduce constraints on the uncontrolled deployment of environmental resources, rather than restricting the dimension of the tourism business. Also, variable return to scale models has mainly been considered, although an analysis using a constant return to scale DEA model has also been conducted on the same data set.

The main result of this DEA performed over the data set is a ranking of the considered municipalities. The analysis for each municipality specifies not only the relative efficiency scores, but also the potential improvements in the case of scores lower than one.

Consumer Durables

A wide program 'Measurement of Eco-efficiency using Data Envelopment Analysis' has been introduced in Finland by the Ministry of the Environment. Economic activity consumes natural resources as its input, and produces undesirable emissions and waste as its output. The empirical approach with the starting point in classical economic theory, which takes into account the substitution possibilities by estimating the so-called efficient production frontiers, takes to natural understanding that a production enterprise is called efficient if the consumption of any input cannot be decreased without the corresponding decrease of at least one output or increase of undesirable outputs. The objective of this project is to investigate the applicability of the DEA-method to the measurement of eco-efficiency, and develop the method further on towards a more comprehensive framework supporting the management and incentive mechanisms. The paper by (Kuosmanen & Kortelainen, 2004) is a very good introduction for how to use DEA in environmental evaluation and for comparative assessment of firm performance in this context.

In (Kortelainen & Kuosmanen, 2005) a method for eco-efficiency analysis of consumer durables by utilizing the DEA was developed. The novel innovation of the paper is to measure the efficiency in terms of absolute shadow prices that are optimized endogenously within the model to maximize the efficiency of the product and producer. The approach is illustrated by an application to eco-efficiency evaluation of a sort of consumer durables, Sport Utility Vehicles. To assess eco-efficiency of a product, one needs to account for both private net economic benefits and external social costs that arise during the use phase of the product's life-cycle. The authors mention that DEA seeks endogenously determined optimal, the so-called shadow prices that present every consumer durable in the most favorable light when compared to other products, and that the method does not require any prior arbitrary assumption as on how to set the prices of environmental pressures.

The key idea of the approach is to test whether there are any nonnegative efficiency prices at which a consumer durable is efficient. The definition of socially efficient is that a product needs to fulfill the conditions of inactivity and optimality. Firstly, it means that the value added for the consumer durable has to be nonnegative at optimal prices and the rationale behind the condition is that the consumers can be inactive, and not purchase any of the goods if the costs outweigh the benefits. The optimality condition demands that the consumer durable must be the optimal choice at some efficiency prices. The goods are eco-efficient if the shadow price of at least one environmental pressure must be positive. Using the efficiency measures and the shadow prices, all goods are classified into the following categories: efficient goods; eco-efficient goods; weakly efficient, economical goods; inefficient goods; inefficient, but environmentally friendly goods; inefficient, environmentally harmful goods.

The authors calculated efficiency scores for 88 different models of Sport Utility Vehicles by using the absolute shadow price approach. For the comparison, the efficiency scores with the environmental efficiency DEA approach were estimated where environmental pressures were modeled as negative outputs. The fuel costs and all environmental pressures were measured per 1 kilometer, which was simultaneously the value of (desirable) output, so the DEA model was invariant to the returns to scale (RTS) specification; all alternative RTS specifications yielded exactly the same results.

Irrigation Districts Management

Application of the DEA is proposed as a methodology to assign the correct weightings for the calculation of indexes and to overcome the subjectivity of the interpretations of the results in management of Andalusian irrigation districts (Spain). The case was presented and discussed by (Rodríguez Díaz et al., 2004). This study was used to select the most representative irrigation districts in Andalusia which were then studied in greater depth. Andalusia is a region of southern Spain, a typical Mediterranean region where irrigation and wealth have been closely linked over time and where 815,000 hectares of irrigated area is divided into 156 irrigation districts. In addition to allowing the production of winter crops, irrigation makes it possible to produce a larger number of crops during the extremely dry summers that are characteristic of this Mediterranean climate, something that would otherwise not be possible under dry-land agriculture. The input-oriented DEA model was used and applied for all the irrigation districts, and separately for interior districts. The authors show that two types of clearly differentiated agricultural districts coexist in Andalusia: interior and littoral districts.

In this research input parameters for DEA application were: irrigated surface area in hectares, labor in annual working units and the total volume of water applied to an irrigation district. The total value of agricultural production in Euros was considered as the output parameter. Following the DEA, none of the districts in the interior achieved high efficiencies in numerical experiments where all districts were considered. This leads to the conclusion that the littoral districts would serve as the reference region for districts in the interior. For this reason, the DEA model was applied separately to the interior districts only, and then it was possible to point out important conclusions. Particularly, the DEA study allowed the five most representative irrigation districts of Andalusia to be selected for a more detailed benchmarking study and the DEA turned out to be a useful tool for detecting local inefficiencies and determining the possible improvements for irrigation.

Biogas Plants

Today it is widely recognized that the largest source of atmospheric pollution is fossil fuel combustion, which current energy production and use patterns rely heavily on, and therefore, the most crucial environmental problems derive from the energy demand to sustain human needs and economic growth. In the paper (Madlener et al., 2006) we find an interesting study of assessing the efficiency of 41 agricultural biogas plants in Austria. The two input parameters were: the amount of organic dry substrate used and the labor time spent for plant operation. Among the three outputs two were desirable: the net electricity produced and external heat which correspondingly refer to the amount of electricity and heat delivered by the biogas plant for external consumption (i.e. net of what the biogas plant consumes itself), including farm operations not directly related to the biogas plant. The third output parameter, methane emissions to the atmosphere, was an undesirable output that contributes to the greenhouse gas problem. In the paper one can find a detailed discussion of DEA efficiency interpretation for biogas plants under consideration.

National Park Management

Wilderness protection is another growing necessity for modern societies, particularly for areas where population density is extremely high and where during the twentieth century the erosion of territory, hence of ecosystems, due to human activities, was dramatically increasing, as for example in Europe. Conservation, however, implies very high opportunity costs and thus it is crucial to create incentives for efficient management practices, to promote benchmarking and to improve conservation management. A methodology based on DEA for assessing the relative efficiency of the management units of the protected area and to indicate how it could be improved, is proposed in (Bosetti & Locatelli, 2005). In it 17 Italian National Parks (National Park Management offices) are considered as DMUs.

Three different models have been used to perform DEA. They differ by the choice of input and output indicators. The set of input parameters contains economic costs, computed aggregating management costs and variable costs and extraordinary expenses. The area extension was considered as a proxy of fixed costs, which were assumed to be proportional to the area covered by the park. The output parameters are: the number of visitors to the park as an indicator of its attractiveness, providing potential indirect benefit to the local economy; the number of the parks' employees, as an indicator of the social and economic indirect and direct benefits; the number of economic businesses which are directly linked, empowered or created thanks to the presence of the park, e.g. the farmers producing within the protected area; the number of protected species, as a good proxy of the environmental quality and biodiversity of the park; and the number of students who visit the park for environmental education trips, as a proxy of social and educational benefits deriving from the park. In some models the inverse of the mentioned biodiversity indicator was included as an input.

There are several different definitions of ecological efficiency known. In the case of protected areas, the problem of efficiency becomes even more complicated, because management and financial features have to be considered as well. In this research this definition appears in the DEA application results: when a DMU is scoring maximum efficiency according to all three models, one can say that the management has attained the sustainable development goal in a very broad sense. In cases where DMUs are partially inefficient the authors use the DEA to obtain information concerning potential improvements in the management. In the final conclusions it is pointed out that the DEA is a good benchmarking technique to monitor multi-objectives efficiency. The DEA also provides the possibility for a detailed analysis of potential improvements in the management of National Parks.

Measuring Residential Energy Efficiency

In the paper (Grösche, 2009) the energy efficiency improvements of US single-family homes between 1997 and 2001, using a two-stage procedure, is estimated. In the first stage, an indicator of energy efficiency is derived by means of Data Envelopment Analysis, and the analogy between the DEA estimator and traditional measures of energy efficiency is demonstrated. The second stage employs a bootstrapped truncated regression technique to decompose the variation in the obtained efficiency estimates into a climatic component and factors attributed to efficiency improvements. The author claims that the improvement of residential energy efficiency is one major goal of energy policy makers.

Put simply, DEA can be considered as a generalization of the energy efficiency defined as the ratio between the amount of a particular produced service **s** and the amount of energy **e** consumed for the production (**s**/**e**). The households' total energy consumption serves as the only input for DEA, measured in kWh. Considering the outputs (the 'produced' energy services), the author approximates the demand for space heating and cooling, and lightning with the size of living space. The number of household members serves as a proxy for the amount of hot water preparation and cooked meals. To account for energy consumption due to the use of electric appliances, the joint number of TV-sets, videos, DVDs, and computers is incorporated. The overall number of refrigerators and freezers in the household is likewise included in the estimation.

The results of the study are mixed: a substantial part of the variation in efficiency scores is due to climatic influences, but households have nevertheless improved their energy efficiency. In particular, households heating mainly with fuel oil or natural gas show significant improvements. A key advantage of the applied procedure is its ease in measuring residential energy efficiency improvements.

Pig Fattening Farms

In the paper (Lauwers & van Huylenbroeck, 2003) a method for analyzing environmental efficiency of Belgian pig fattening farms, based on the farm materials balance and worked out, the DEA technique is proposed. In its most fundamental sense, materials balance means the mass flow equation of raw materials used in the economic system, and of the residuals disposed of in the natural environment. The environmental efficiency of the farm is defined similarly to the economic allocative efficiency of farms as DMUs in the DEA approach. Nutrient surplus in pig fattening, as a typical balance indicator, is used to illustrate the concept in a two input – one output case. The input parameters are feed (nutrients) and the piglets or rotation. Because of juridical constraints on farm dimension, the farmers profit maximization objective turns into a maximization of gross margin per pig place, thus the pig fattening process is simplified to one output, the marketable meat production. Several versions of DEA are applied and discussed: input-oriented and output-oriented. The main conclusion is that, ignoring the balance feature of environmental issues, such as nutrient surplus, might be the main reason why traditional integral analyses of economic and environmental efficiency yield contradictory conclusions.

Environmentally Friendly Manufacturing

The essence of environmentally friendly manufacturing is to define sustainable development in terms of manufacturing; conserving nature's services (resource supplier, water, energy) with the development being centered on economics and trade with a timeline of 20-100 years. Today the corporate environment is an evolving strategy, and it is not surprising that manufacturing operations are required to consider environmental impacts and sustainable development more and more. In Wu et al. (2006) the application of DEA is discussed to measure the efficiency, through material loss and environmental impact of a closed-loop manufacturing process, of a process in the computer industry. The multi-stage DEA is being utilized to measure each manufacturing phase, to compare the starting materials and environmental status to the materials available once the product has been recycled, and the environmental damage (or substances added to the environment) once a product lifecycle has been completed. The multi-stage DEA becomes important for multiple processes, for instance, the closed-loop manufacturing process, and in this case, the outputs from one process can be the inputs for the next. Inputs and outputs varied, the whole set of parameters consisted of expenditure on research and development, years of experience, expenditures on raw materials, number of product parts, use of harmful materials, energy consumption, product recyclability in percents of product/components reusable, emission of pollutants, modification flexibility, amount of products returned, product recyclability and material recovered. The values for the various inputs/outputs were obtained from the literature. This work is a good example of measuring a company's environmental conduct in manufacturing by applying the DEA. The conclusion is that if one organization has multiple locations or is responsible for multiple products, the multi-stage DEA is a promising approach. In addition, on the base of the model it is possible to analyze a company's environmental impact and motivate the company to improve their corresponding indicator numbers.

Grain Farms

The aim of the study (Vasiliev et al., 2008) was to analyze the efficiency of Estonian grain farms after Estonia's transition to market economy and during the accession period to the European Union in 2000–2004. Here DEA is used with the following input parameters: changing expenditures, total investment capital, the surface of land used and the yearly working units. The output was fixed as total production volume in monetary value. The results obtained were: the mean total technical efficiency varied from 0.70 to 0.78, and 62% of the grain farms are operating under the increasing returns to scale. The most pure technically efficient farms were the smallest and the largest, but the productivity of small farms was low when compared to larger farms because of their small scale. Although, solely based on the DEA model, it is not possible to determine the optimum farm scale and the range of Estonian farm sizes operating efficiently.

Final Remarks

Environmental Technologies mean cleaner and resource efficient technologies which can decrease material inputs, reduce energy consumption and emissions, discover valuable by-products and minimize waste disposal problems or some combination of these. Data Envelopment Analysis is widely used in many areas for estimating the peer-based efficiency of Decision Making Units which are defined in each single case. Although not the only method of quantitative analysis and efficiency modeling, the DEA is highly recommended as an approach with interpretable output, and which has improved its effectiveness as a productivity analysis tool. The primary advantages of this technique are that it considers multiple input and output factors of DMUs and does not require parametric assumptions of traditional multivariate methods. In general, inputs can include any resources utilized by a DMU, and the outputs can range from actual products produced to a range of performance and activity measures. The DEA has several versions and modifications, and each of these models and methods can be useful in a variety of manufacturing and service areas.

REFERENCES

Bosetti, V., & Buchner, B. (2008). Using DEA to assess the Relative Efficiency of Different Climate Policy Portfolio. *Ecological Economics*, *68*(5), 1340–1354. doi:10.1016/j.ecolecon.2008.09.007

Bosetti, V., Lanza, A., & Cassinelli, M. (2004). *Using Data Envelopment Analysis to Evaluate Environmentally Conscious Tourism Management*. FEEM Working Paper 59. Milan: Fondazione Eni Enrico Mattei.

Bosetti, V., & Locatelli, G. (2005). *A Data Envelopment Analysis Approach to the Assessment of Natural Parks' Economic Efficiency and Sustainability*. The Case of Italian National Parks (May 2005). FEEM Working Paper No. 63.05. Available at SSRN: http://ssrn.com/abstract=718621. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=718621

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (Eds.). (1994). *Data envelopment analysis: Theory, methodology, and applications*. Boston: Kluwer.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. doi:10.1016/0377-2217(78)90138-8

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-solver Software* (2nd ed.). New York: Springer.

Cooper, W. W., Seiford, L. M., & Zhu, J. (Eds.). (2004). *Handbook on Data Envelopment Analysis* (2nd ed.). Boston: Kluwer.

Emrouznejad, A. (1995-2001). *Ali Emrouznejad's DEA HomePage*. Warwick Business School, Coventry CV4 7AL, UK, http://www.deazone.com/.

Emrouznejad, A., Parker, B., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Journal of Socio-Economics Planning Science*, *42*(3), 151–157. doi:10.1016/j.seps.2007.07.002

Färe, R., & Grosskopf, S. (2004). Modeling undesirable factors in efficiency evaluation [comment]. *European Journal of Operational Research*, *157*, 242–245. doi:10.1016/S0377-2217(03)00191-7

Grösche, P. (2009). Measuring residential energy efficiency improvements with DEA. *Journal of Productivity Analysis*, *31*(2), 87–94. doi:10.1007/s11123-008-0121-7

Kao, C., & Tai Liu, S. (2000). Data Envelopment Analysis with Missing Data: an Application to University Libraries in Taiwan. *The Journal of the Operational Research Society*, *51*, 897–905.

Kortelainen, M., & Kuosmanen, T. (2005). *Eco-Efficiency Analysis of Consumer Durables Using Absolute Shadow Prices*. EconWPA Working Paper at WUSTL, No. 0511022.

Kuosmanen, T., & Kortelainen, M. (2004). *Data Envelopment Analysis in Environmental Valuation: Environmental Performance, Eco-efficiency and Cost-Benefit Analysis*. Discussion Paper No.21, Department of Business and Economics, University of Joensuu.

Lauwers, L. H., & van Huylenbroeck, G. (2003). Materials Balance Based Modelling Of Environmental Efficiency. In *2003 Annual Meeting, August 16-22, 2003, Durban, South Africa 25916*, International Association of Agricultural Economists.

Lee, S. K., Mogi, G., Shin, S. C., & Kim, J. W. (2008) Measuring the Relative Efficiency of Greenhouse Gas Technologies: An AHP/DEA Hybrid Model Approach. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* 2008 Vol. II, IMECS 2008, 19-21 March, 2008, Hong Kong, (pp. 1615-1619).

Madlener, R., Antunes, C. H., & Dias Luis, C. (2006). *Multi-Criteria versus Data Envelopment Analysis for Assessing the Performance of Biogas Plants*. CEPE Working Paper No.49, Centre for Energy Policy and Economics (CEPE), Zurich.

Miidla, P., & Kikas, K. (2009). The efficiency of Estonian central public libraries. *Performance Measurement and Metrics*, *10*(1), 49–58. doi:10.1108/14678040910949684

Piot-Lepetit, I., Vermersch, D., & Weaver, R., D. (1997). Agriculture's environmental externalities: DEA evidence for French agriculture. *Applied Economics*, *29*(3), 331–338. doi:10.1080/000368497327100

Rodrigues Diaz, J. A., Camacho Poyato, E., & Lopez Luque, R. (2004). Applying Benchmarking and Data Envelopment Analysis (DEA) Techniques to Irrigation Districs in Spain. *Irrigation and Drainage*, *53*, 135–143. doi:10.1002/ird.128

Seiford, L., M., Zhu, J. (2005). A response to comments on modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *161*(2), 579–581. doi:10.1016/j.ejor.2003.09.018

Taniguchi, M., Akinaga, J., & Abe, H. (2000). Evaluation for Neighboring Environment considering Comparative Stduy and DEA Analysis. *Infrastructure Planning Review*, *17*, 423–430.

Thanassoulis, E. (2003). *Introduction to the Theory and Application of Data Envelopment Analysis: a Foundation Text with Integrated Software*. Norwell, MA: Kluwer Academic Publishers.

Vasiliev, N., Astover, A., Mõtte, M., Noormets, M., Reintam, E., Roostalu, H., & Matveev, E. (2008). Efficiency of Estonian grain farms in 2000-2004. *Agricultural and Food Science*, *17*(1), 31–40. doi:10.2137/145960608784182272

Wu, T., Fowler, J., Callerman, T., & Moorehead, A. (2006) Multi-stage DEA as a Measurement of Progress in Environmentally Benign Manufacturing. In *The 16th International Conference on Flexible Automation and Intelligent Manufacturing*, (pp. 221-228), Limerick, Ireland, June, 2006.

Zhou, P., & Anga, B., W. & Poh, K.L. (2008). Measuring environmental performance under different environmental DEA technologies. *Energy Economics*, *30*(1), 1–14. doi:10.1016/j.eneco.2006.05.001

ADDITIONAL READING

A Data Envelopment Analysis (DEA) Home Page. (1996) http://www.etm.pdx.edu/dea/homedea.html.

Beasley, J. E. (1996) *OR-Notes.* http://people.brunel.ac.uk/~mastjjb/jeb/or/dea.html.

Coelli, T. J., & Rao, D. P. O'Donnell. C. J., Battese, G. E. (2005) *An Introduction to Efficiency and Productivity Analysis*, Springer.

Molinero, C. M., & Woracker, D. (2008) *Data Envelopment Analysis a Non-Mathematical Introduction*. Available at SSRN: http://ssrn.com/abstract=6317 Working paper.

Thore, S. A. (Ed.). (2002) *Technology Commercialization: DEA and Related Analytical Methods for Evaluating the Use and Implementation of Technical Innovation*. Springer, 2002.

Zhu, J. (Developer) *DEAFrontier*. http://www.deafrontier.net/.

KEY TERMS AND DEFINITIONS

Constant Returns to Scale: The outputs of a DMU change in the same proportion as inputs; the producers are able to linearly scale the inputs and outputs without increasing or decreasing efficiency.

Decision Making Unit: Decision making units form a homogenous set of peer entities which convert multiple inputs into multiple outputs and efficiency of which is under consideration in DEA.

Efficiency (Computed in Data Envelopment Analysis): Weighted sum of outputs divided by the weighted sum of inputs.

Environmental Technologies: Cleaner and resource efficient technologies which can decrease material inputs, reduce energy consumption and emissions, discover valuable by-products, and minimize waste disposal problems or some combination of these.

Relative Efficiency: A decision making unit is to be rated as fully efficient on the basis of available evidence if and only if the performances of other decision making units do not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs.

Variable Returns to Scale: The outputs of a decision making unit change in other proportion as inputs, increasingly or decreasingly.

This work was previously published in Environmental Modeling for Sustainable Regional Development: System Approaches and Advanced Methods, edited by Vladimír Olej, Ilona Obršálová and Jirí Krupka, pp. 242-259, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 37
Constrained Optimization of JIT Manufacturing Systems with Hybrid Genetic Algorithm

Alexandros Xanthopoulos
Democritus University of Thrace, Greece

Dimitrios E. Koulouriotis
Democritus University of Thrace, Greece

ABSTRACT

This research explores the use of a hybrid genetic algorithm in a constrained optimization problem with stochastic objective function. The underlying problem is the optimization of a class of JIT manufacturing systems. The approach investigated here is to interface a simulation model of the system with a hybrid optimization technique which combines a genetic algorithm with a local search procedure. As a constraint handling technique we use penalty functions, namely a "death penalty" function and an exponential penalty function. The performance of the proposed optimization scheme is illustrated via a simulation scenario involving a stochastic demand process satisfied by a five–stage production/inventory system with unreliable workstations and stochastic service times. The chapter concludes with a discussion on the sensitivity of the objective function in respect of the arrival rate, the service rates and the decision variable vector.

INTRODUCTION

This chapter addresses the problem of production coordination in serial manufacturing lines which consist of a number of unreliable machines linked with intermediate buffers. Production coordination in systems of this type is essentially the control of the material flow that takes place within the system in order to resolve the trade-off between minimizing the holding costs and maintaining a high service rate. A time-honored approach to modeling serial manufacturing lines is to treat them as Markov Processes (Gershwin, 1994, Veatch and Wein, 1992) and then solve the related Markov Decision Problem (MDP), by using standard iterative algorithms such as

DOI: 10.4018/978-1-4666-1945-6.ch037

policy iteration, (Howard, 1960), value iteration, (Bellman, 1957) etc. However the classic dynamic programming (DP), approach entails two major drawbacks: Bellman's curse of dimensionality, i.e. the computational explosion that takes place with the increase of the system state space, and the need for a complete mathematical model of the underlying problem. The limitations of the DP approach gave rise to the development of sub-optimal yet efficient production control mechanisms.

A class of production control mechanisms that implement the JIT (Just In Time) manufacturing philosophy known as *pull type* control policies/mechanisms has come to be widely recognized as capable of achieving quite satisfactory results in serial manufacturing line management. Pull type control policies coordinate the production activities in a serial line based only on actual occurrences of demand rather than demand forecasts and production plans as is the case in MRP-based systems. In this chapter, six important pull control policies are examined, namely Kanban and Base Stock (Buzacott and Shanthikumar, 1993), Generalised Kanban (see Buzacott and Shanthikumar (1992), for example), Extended Kanban (Dallery and Liberopoulos, 2000), CONWIP (Spearman *et al*., 1990) and CONWIP/Kanban Hybrid (Paternina-Arboleda and Das, 2001). Pull production control policies are heuristics characterised by a small number of control parameters that assume integer values. Parameter selection significantly affects the performance of a system operating under a certain pull control policy and is therefore a fundamental issue in the design of a pull-type manufacturing system. In this chapter the performance of JIT manufacturing systems is evaluated by means of discrete-event simulation (Law and Kelton, 1991). In order to optimize the control parameters of the system the simulation model is interfaced with a hybrid optimization technique which combines a genetic algorithm with a local search procedure.

The application of simulation together with optimization meta-heuristics for the modeling and design of manufacturing systems is an approach that has attracted considerable attention over the past years. In Dengiz and Alabas (2000) simulation is used in conjunction with tabu search in order to determine the optimum parameters of a manufacturing system while Bowden *et al*. (1996) utilize evolutionary programming techniques for the same task. Alabas *et al.* (2002) develop the simulation model of a Kanban system and explore the use of genetic algorithm, simulated annealing and tabu search to determine the number of kanbans. Simulated annealing for optimizing the simulation model of a manufacturing system controlled with kanbans is applied in Shahabudeen *et al.* (2002), whereas Hurrion (1997) constructs a neural network meta-model of a Kanban system using data provided by simulation. Koulouriotis *et al.* (2008) apply Reinforcement Learning methods to derive near-optimal production control policies in a serial manufacturing system and compare the proposed approach to existing pull type policies. Some indicative applications of genetic algorithms (GAs) in manufacturing problems can be found in Yang *et al.* (2007), Yamamoto *et al* (2008), Smith and Smith (2002), Shahabudeen and Krishnaiah (1999) and Koulouriotis *et al.* (2010). Panayiotou and Cassandrass (1999) develop a simulation-based algorithm for optimizing the number of kanbans and carry out a sensitivity investigation by using finite perturbation analysis. It has been suggested in the literature that the results of a genetic algorithm can be enhanced by conducting a local search around the best solutions found by the GA, (for related work see Yuan, He and Leng, 2008 and Vivo-Truyols, Torres-Lapasio and Garcıa-Alvarez-Coque, 2001). On that basis, this hybrid optimization scheme has been adopted in the present study.

The main contributions of this work are the following. The performance of six important pull production control policies in a hypothetical scenario is investigated using discrete event simulation. In order to determine the control parameters of each policy the proposed hybrid

GA is employed. The objective function to be optimized is a weighted sum of the mean Work In Process, (WIP), inventories subject to the constraint of maintaining the service level, (*SL*), above a specified target. Due to the fact that the objective function is stochastic we use resampling, i.e., performing multiple evaluations of the same parameter vector and using the mean of these evaluations as the fitness measurement of this individual, a practice discussed by Fitzpatrick and Grefenstette, (1988) and Hammel and Bäck, (1994). As a constraint handling technique two types of penalty functions are explored; a "death penalty" function and an exponential penalty function. The exponential penalty function is designed according to an empirical method which is based on locating points which lie on the boundaries between feasible and infeasible region from the output of the genetic algorithm with the "death penalty" function. Our numerical results support the intuitive perception that the "death penalty" approach most of the times will yield worse results than the exponential penalty function which penalizes solutions according to the level of the constraint violation. The chapter concludes with a discussion on how the objective function behaves for different levels of arrival rate and service rates as well as on its sensitivity to the decision variable vector.

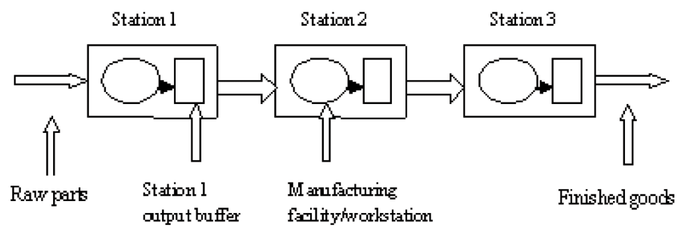
The remaining material of this chapter is structured as follows. Sections "Base Stock Control Policy" to "CONWIP/Kanban Hybrid Control Policy" give a brief description of six important pull production control policies for serial manufacturing lines. In sections "Optimization Problem: Objective Function" and "Hybrid Genetic Algorithm" we discuss the main aspects of the simulation optimization methodology that we followed and namely, the formal definition of the parameter optimization problem and issues concerning the genetic algorithm and local search procedure that was used. We report our findings from the simulation experiments that we conducted for one serial line starting from section "Experimental Results: Simulation Case" and thereafter. Finally, in the last section we state our concluding remarks and point to possible directions for future research.

SYSTEM DESCRIPTION: JIT PRODUCTION CONTROL POLICIES

We examined manufacturing serial lines that produce a single product type and consist of a number of workstations/machines with intermediate buffers. We assume that the first machine is never starved. Customer demands arrive at random time intervals and request the release of one finished part from the finished goods buffer. Demands are satisfied immediately from the finished parts inventory while in the case where there are no parts available in the last buffer the demand is backordered. We do not consider customer impatience in our model, so no demand is ultimately lost to the system. Manufacturing facilities have the ability to work on only one part at a time during a manufacturing cycle. All machines have random production time, time between failures and repair time. As soon as a stage *i* part is manufactured, it is placed in the output buffer of that station. A control policy coordinates the release of parts from the output buffer of that station to the next machine. The unreliability of the manufacturing operations along with the stochastic demand for final products dictates the use of safety buffers of intermediate and finished parts in order to attain the target service rate. However the use of safety stocks incurs significant holding costs that could bring the manufacturer to a position of competitive disadvantage, and therefore, it is essential to balance the trade-off between minimizing WIP inventories and maintaining a high service level. Figure 1 shows a manufacturing system with three stations in tandem.

The following sections briefly explain the way that the Kanban, Base Stock, CONWIP, CONWIP/Kanban Hybrid, Extended Kanban and Generalized Kanban control policies for serial lines operate.

Figure 1. A three station manufacturing line



BASE STOCK CONTROL POLICY

A Base Stock, (see Buzacott and Shanthikumar, 1993), manufacturing line is completely described by *N* parameters, the base stock levels S_i of each production station, $i = 1, 2, \ldots, N$ where *N* is the number of the system's workstations. The S_i parameters correspond to the number of parts that exist in the system's buffers at the time the system is in its initial state that is before any demands have arrived to the system. This control policy operates as follows. When a demand arrives to the system it is immediately transmitted to every manufacturing station, authorizing it to start working on a new station *i* part. Base Stock has the advantage of reacting rapidly to incoming demand, with the drawback of providing no control at all on the system's inventories.

KANBAN CONTROL POLICY

The Kanban control policy that was originally developed by the Toyota Motor industry and became the topic of considerable research thereafter, (Sugimori et al, 1977, Buzacott and Shanthikumar, 1993, Berkley, 1992, Karaesmen and Dallery, 2000). A Kanban manufacturing line's control parameters are the production authorizations K_i of each station, $i = 1, 2, \ldots, N$. The K_i parameter corresponds to the maximum number of parts that are allowed in station *i* (manufacturing facility – output buffer). Workstation *i* is authorized to start working on a new part as soon as a finished station *i* part is released from its output buffer. The information of a demand arrival is transmitted from the last manufacturing station to the first one station – by – station. If there is a buffer with no parts in it then this transmission is interrupted. The Kanban policy offers very tight synchronization between the various production stations of the system at the expense of the relatively slow response to demand fluctuations.

CONWIP CONTROL POLICY

CONWIP is an abbreviation for CONstand Work In Process (Spearman et al, 1990). According to this policy the total number of parts that exist in the system, (Work In Process), can never exceed a certain level, which is the *C* control parameter of the policy. Parameter *C* is equal to the sum of the system's base stocks S_i, $i = 1, 2, \ldots, N$. All machines in a CONWIP line are authorized to produce whenever they have this ability, (they are operational and have a raw part to work on), except the first one. The first machine of the system is authorized to start working on a new part as soon as a unit from the finished parts buffer is released to a customer.

GENERALIZED KANBAN AND EXTENDED KANBAN CONTROL POLICIES

These two control policies combine the merits of Base Stock and Kanban as they react rapidly to the arrival of demands and effectively control the WIP at the same time. They are described by two parameters per station, the base stocks S_i and the production authorizations K_i,($K_i \geq S_i$), which are borrowed from the Base Stock and Kanban policies respectively. The finite number of production authorizations guarantees that the system's inventories will not exceed the pre – defined levels, but the station coordination here is not as tight as in Kanban. A station can be granted a production authorization even if a part is not released from its output buffer. For a detailed description of the way Generalized Kanban and Extended Kanban operate the reader is referred to Liberopoulos and Dallery (2000) and Buzacott and Shanthikumar (1992).

CONWIP/KANBAN HYBRID CONTROL POLICY

A CONWIP/Kanban Hybrid system (see Paternina – Arboleda and Das (2001) for example), as implied, operates under a combination of the CONWIP and Kanban control policies. Departure of a finished part from the system authorizes the first station to allow a new raw part to enter the system. All workstations except the last one have a finite number of production authorizations K_i, $i = 1, 2, \ldots, N-1$. Station production authorizations K_i, the base stock S_N of the last workstation and the total WIP, (parameter C), that is allowed in the system are CONWIP/Kanban Hybrid's control parameters.

OPTIMIZATION PROBLEM: OBJECTIVE FUNCTION

The mathematical formulation of the parameter optimization problem for serial lines controlled by pull production control policies is given below. Let $\mathbf{x} = [x_1\ x_2 \ldots x_n]$, $x_i \in \mathbf{Z}$, be the control parameter vector of some pull production control policy, i.e. the station i production authorizations (kanbans) in a Kanban system, or the initial buffer levels in a Base Stock system etc. The objective is to find the control parameter values $\mathbf{x}$ that maximize the expected value of the stochastic objective function $f(\mathbf{x}, \omega)$, subject to the constraint of maintaining the service level (*SL*) equal to or above a specified target t.

maximize:

$$E\left[f\left(\mathbf{x}, \omega\right)\right],\ x_i \in \mathbf{Z},\ i = 1, 2, \ldots., n \tag{1}$$

subject to:

$$E\left[SL\left(\mathbf{x}, \omega\right)\right] \geq t \tag{2}$$

ω is used to denote the stochastic nature of f, and *SL*. *SL* is an unknown function of $\mathbf{x}$ and $t \in \Re^{+}$. The evaluation of functions f and *SL* is the result of a simulation experiment. The value of *SL* is the number of demands satisfied by on-hand inventory divided by the number of total demands which arrived to the system. The value of f is a weighted sum of the mean WorkInProcess inventories and is calculated according to 3.

$$f = -\sum_{i=1}^{N} h_i \bar{H}_i \tag{3}$$

where h_i stands for the cost of storing one item in output buffer i per time unit, and $\overline{H}_i$ is the average inventory level in buffer i We know that the optimal solution in this type of problems is located very close to the boundaries between feasible and infeasible region. Additional difficulty in obtaining the optimal solution emanates from the fact that fitness measurements contain random "noise" caused by the simulation model.

HYBRID GENETIC ALGORITHM

In order to solve the optimization problem stated in the previous section we propose a hybrid optimization technique which combines a genetic algorithm with a local search procedure. The genetic algorithm evolves a population of candidate solutions (individuals), where each solution is evaluated with the use of a simulation model, and the individual with the highest fitness value found by the GA is used to initialize the local search procedure. The fitness $v\left(\mathbf{x}\right)$ of each individual is represented by Equation (4).

$$v\left(\mathbf{x}\right) = \frac{1}{m}\sum_{i=1}^{m}\left[f(\mathbf{x},\omega) + p\left(SL(\mathbf{x},\omega)\right)\right] \quad (4)$$

where $f(\mathbf{x},\omega)$ is calculated according to (3), $p\left(\mathbf{x}\right)$ is a properly defined penalty function of the service level, and *m* is a positive integer (sample size). The parameters of the GA are the chromosome length *l*, the population size *s*, the sample size m, a positive integer *e* called elite count, the crossover probability P_{cross}, the mutation probability P_{mut} and the max number of generations *g*. The individuals which constitute the genetic algorithm population are encoded as binary bit-strings, therefore parameter *l* controls the size of the search space. Parameter *e* determines the number of individuals that pass deterministically to the next generation. The local search procedure is characterized by a single parameter $\delta \in \Re^{+}$. Let $\mathbf{x}_{cur}$ be the current solution of the local search algorithm, v_{cur} its fitness value and v_{best} the best fitness value found so far. If we denote the search space by S and a distance function, (e.g. Euclidean distance), by *dist()*, then the neighborhood of $\mathbf{x}_{cur}$ is written as $N\left(\mathbf{x}_{cur}\right) = \left\{\mathbf{y} \in S : dist(\mathbf{x},\mathbf{y}) \le \delta\right\}$. The pseudocode of the hybrid genetic algorithm is presented below.

1. Input GA parameters: chromosome length *l*, population size *s*, sample size m, elite count *e*, crossover probability P_{cross}, mutation probability P_{mut}, max number of generations *g*
2. Initialize population randomly, set *generation_counter*←0
3. WHILE(*generation_counter* $< g$)
   a. evaluate population. set

$$generation\_counter \leftarrow generation\_counter + 1$$

   b. scale fitness values proportionally to raw fitness measurements
   c. apply selection operator
      - select *e* individuals with highest fitness values
      - select the remaining $s - e$ individuals using stochastic uniform selection
   d. apply crossover operator
   e. apply mutation operator
4. return individual $\mathbf{x}_{best}$ with highest fitness
5. Initialize local search algorithm: $\mathbf{x}_{cur} \leftarrow \mathbf{x}_{best}$, define neighborhood parameter δ
6. evaluate $\mathbf{x}_{cur}$. Set $v_{best} \leftarrow v_{cur}$, $flag \leftarrow TRUE$
7. WHILE $(flag = TRUE)$
   a. evaluate all points in $N(\mathbf{x}_{cur})$
   b. select $\mathbf{x}_{new} \in N(\mathbf{x}_{cur})$ with best fitness value v_{new}
   c. IF $(v_{new} > v_{best})$ THEN set $v_{best} \leftarrow v_{new}$, $\mathbf{x}_{cur} \leftarrow \mathbf{x}_{new}$ ELSE THEN $flag \leftarrow FALSE$
8. return $\mathbf{x}_{cur}$. Terminate

The selection operator (Step 3.c) determines which individuals will be chosen to create the next generation. The first e individuals in terms of fitness value pass to the next generation by default. The remaining s - e individuals are selected with the use of a stochastic uniform selection routine. This technique can be visualized as a line in which each individual corresponds to a section of the line of length proportional to its scaled fitness value. The algorithm moves along the line in equal-sized steps. At each step, the algorithm selects an individual from the corresponding section it finds itself on. In the crossover stage (Step 3.d), pairs of individuals are selected at random with probability $P_{cross} \in (0,1)$ in order to be recombined. In the implementations of the GA for the one-parameter-per-workstation manufacturing systems we used the single-point crossover method. For the remaining systems, (Extended and Generalized Kanban) uniform crossover was used. According to this technique, two individuals exchange bits on the basis of a randomly generated binary vector of equal length called crossover mask. The mutation operator (Step 3.e) modifies the value of a bit in the population with probability $P_{mut} \in (0,1)$. The genetic algorithm terminates when it completes a predefined number of iterations g and returns the individual with the highest fitness value which is used to initialize the local search algorithm. The complexity of the hill-climbing procedure is $O\left(t \times k\right)$, where *t* is the number of iterations and *k* the neighborhood size. The complexity of the genetic algorithm depends on the number of generations, the size of the population and the genetic operators/parameters used.

EXPERIMENTAL RESULTS: SIMULATION CASE

We examined a five-machine manufacturing line with equal operation times. The base simulation scenario consists of the following parameters: Machines operate with service rates which are normally distributed random variables with mean 1.1 parts/time unit and st.d. 0.01. Repair to failure times are exponentially distributed with mean 1000 time units. Failures are operation dependent. Repair times are also assumed exponential with a MeanTimeToRepair of 10 time units. Times between two successive customer arrivals are exponential random variables with mean 1.11 time units, i.e. the arrival rate is $R_a = 0.9$. Since the service rates are all equal to 1.1 parts/ time unit and the machines are failure-prone, the maximum attainable throughput rate under any control policy will be $T_{\max} < 1.1$. Consequently, the arrival rate set to 0.9 parts/time unit corresponds to a heavy loading conditions simulation case. The inventory costs for storing one part per time unit in buffer *i* are $\mathbf{h} = [h_1 \ h_2 \ldots h_5]$ = [1.0 1.2 1.44 1.73 2.07]. Note that the holding costs increase at a rate of 20% when moving downstream from buffer to buffer. This increase is due to the value which is added to a part as it is progressively converted into a final product. The system operates under a complete backordering policy, which means that no customer demand is ultimately lost to the system. The justification for selecting the aforementioned probability distributions in order to model the arrival process, the service rates etc. can be found in queueing theory and in manufacturing systems literature. Some indicative references, among others, are the influential works of Law and Kelton (2000) and Bhat (2008). The input parameters of the simulation model were selected in a manner to mimic a situation where the system is under heavy loading conditions. This is a case of primary interest since the differences in performance between the various pull type control policies are most clearly illustrated when the manufacturing line is pushed towards its maximum throughput rate. In order to investigate the sensitivity of the production/inventory system under examination for different levels of arrival rates and service rates as well as the robustness of the solutions obtained by the proposed optimization methodology, four variants of the base simulation scenario are also considered. In the first two variants, all inputs to the simulation model are kept constant except for the arrival rates which are set to $R_a = 1.0$ and $R_a = 0.8$

Table 1. Simulation scenarios parameters

	Ra	*Rp*	*st.d.*	*MTBF*	*MTTR*	*inventory costs*
base case	0.9	1.1	0.01	1000	10	**h**
variant 1	1.0	1.1	0.01	1000	10	**h**
variant 2	0.8	1.1	0.01	1000	10	**h**
variant 3	0.9	1.1	0.1	1000	10	**h**
variant 4	0.9	1.1	0.001	1000	10	**h**

respectively, i.e. we examine the system's behaviour for increased/decreased demand for final products. In the remaining two variants of the base simulation case we vary the standard deviation of the service rates. In one case the system's performance is evaluated for service rates that vary significantly around the mean (st.d. = 0.1) and in the other case we examine what would happen if the "randomness" of the service rates decreased (st.d. = 0.001). The system configuration for the five simulation cases is presented in Table 1.

The goal is to maximize the expected value of the weighted sum of the mean WorkInProcess inventories $f = -\sum_{i=1}^{N} h_i \bar{H}_i,\ i = 1,2,..,5$, subject to the constraint of $E\left[SL(\mathbf{x})\right] \geq 90.0\%$.

HYBRID GENETIC ALGORITHM PARAMETERS

The dimensions of the related optimization problem for the Kanban, Base Stock, CONWIP and Kanban/CONWIP Hybrid systems are *dim* = 5. For the Extended and Generalized Kanban systems the dimensionality of the problem rises to *dim'* = 10. The authors conducted a series of pilot experiments in order to come up with the most suitable hybrid genetic algorithm parameters for this particular problem. An important issue was to resolve the trade-off between quality of final solution and computational cost as the evaluation of the fitness value of the candidate solutions is computationally expensive. We experimented with population sizes in the range [20,50], crossover probabilities in the range [0.3, 0.8] and mutation probabilities in the range [0.001, 0.1]. For the one-parameter-per stage policies the single-point crossover operator was implemented whereas for the two-parameter-per-stage policies we applied uniform crossover. The reason for making this distinction is that offspring produced with the latter crossover technique are generally more diverse compared to their parents than offspring generated by single-point crossover. This is a desirable property due to the fact that the search space for the two-parameter-per-stage policies is by orders of magnitude larger that the search space for the one-parameter-per-stage policies and therefore an intense exploration strategy is required. The neighborhood of the local search algorithm was set to include all data points around the current point **x** with Euclidean distance equal to or less than 1: $N\left(\mathbf{x}\right) = \left\{\mathbf{y} \in S : \sqrt{\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \leq 1\right\}$. Given that the decision variables are integers this is obviously the minimum neighborhood size one could select but since the major part of the search is carried out by the genetic algorithm and the local search procedure is used merely to fine-tune the already obtained solutions, it is acceptable to use a small neighborhood. Admittedly, the parameters of the optimization algorithm were initialized heuristically and one cannot discard the possibility that different values for the parameters could yield better results but a full factorial experiment for the design of the optimization scheme would fall beyond the scope of this chapter. The genetic algorithm's parameters that were ultimately selected are: population size = 30, crossover and mutation probabilities, $P_{cross} = 0.5$ and $P_{mut} = 0.05$ respectively. The individual which scored the highest fitness value passes to the next generation with probability 1, i.e. the elite count parameter was set to 1. Each individual was evaluated 50 times, *m* = 50, where each replicate was executed for 80,000 time units. The GA produced 100 generations for the problems with dimensionality *dim* = 5, (optimizing Kanban, Base Stock, CONWIP, CONWIP/Kanban Hybrid systems). For the problems with dimensionality *dim'* = 10, (optimization of Extended Kanban and Generalized Kanban systems) the GA produced 240 generations of individuals.

COMPUTATIONAL COST

The simulators for the six pull type manufacturing systems as well as the proposed optimization algorithm were coded in C++ and the experiments were conducted on a PC with AMD Athlon processor at 1.8 GHz and 512 MB RAM. The factor that primarily affects the execution time of the hybrid GA is the control parameter evaluation, i.e. the computational cost of the simulation model. Every solution evaluation, that is 50 independently seeded executions of the simulation model, lasts approximately 5 seconds and therefore the evaluation of a generation of candidate solutions (30 individuals) takes about 2.5 minutes to complete. The execution of the hybrid GA for a one-parameter-per-stage policy (100 generations) lasts approximately 4.7 hours, of which 4.2 hours are consumed by the simulation model. On the other hand, the execution of the hybrid GA for a two-parameter-per-stage policy (240 generations) lasts approximately 11 hours, where 10 hours are devoted to the solution evaluation phase.

DEATH PENALTY RESULTS

For the implementation of the hybrid genetic algorithm with "death" penalty we used the following penalty function.

$$p\left(\mathbf{x}\right) = \begin{cases} 0.0, & if \quad E\left[SL(\mathbf{x})\right] \geq 90.0\% \\ -1000.0, & if \quad E\left[SL(\mathbf{x})\right] < 90.0\% \end{cases} \tag{5}$$

We reiterate that the expected value $E\left[SL(\mathbf{x})\right]$ is the arithmetic mean of *m* measurements of *SL*. This is a very straight-forward implementation. Every individual that does not satisfy the service level constraint is penalized heavily and will be probably discarded in the next iteration of the algorithm. The results from the hybrid genetic algorithm runs for each control policy are displayed in Table 2. The rows containing the results of the standard genetic algorithm, (without the local search component), are labeled with the initials GA followed by the control policy's name, while the results of the hybrid algorithm are labeled with the initials GAL and the control policy's name. We made this distinction in order to clarify whether the local search offers some significant improvement or not. The last column of Table 2 contains the fitness values calculated according to Equation (4) of the corresponding parameter sets.

The CONWIP system scored the highest fitness value $v_b = -25.29$ followed by the Generalized Kanban, the Extended Kanban, the Hybrid CONWIP/Kanban, the Kanban and the Base Stock systems in decreasing fitness value order. With the exception of the CONWIP system, the local search algorithm enhanced the best fitness values found by the standard genetic algorithm in a range from 1.35% to 9.34%. It is important to stress here that this improvement refers to the fitness values and not the actual objective function values of the optimization problem. In the case of the Generalized Kanban system the local search al-

Table 2. Best parameter sets and fitness values for "death" penalty function (n.i. stands for "no improvement")

Policies	*x1(/x1')*	*x2(/x2')*	*x3(/x3')*	*X4(/x4')*	*x5(/x5')*	*C*	*v(x)*
GA_Kanban (K_i)	2	2	1	8	13	-	-32.11
GAL_Kanban (K_i)	1	2	1	8	12	-	-29.12
GA_BaseStock (S_i)	2	7	2	0	14	-	-31.70
GAL_BaseStock (S_i)	1	6	2	0	14	-	-29.49
GA_CONWIP (S_i, C)	0	5	3	5	6	19	-25.29
GAL_CONWIP (S_i, C)	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
GA_Hybrid (K_i, i=1,2,3,4, B_5, C)	1	1	3	7	10	22	-28.75
GAL_Hybrid(K_i,i=1,2,3,4, S_5, C)	1	1	2	7	10	21	-26.71
GA_E. Kanban (K_i/S_i)	10/0	11/2	9/1	2/2	23/15	-	-26.54
GAL_E. Kanban (K_i/S_i)	6/0	11/2	6/1	2/2	23/15	-	-26.18
GA_G. Kanban (K_i/S_i)	8/4	2/0	15/3	16/2	14/14	-	-1028.14
GAL_G.Kanban (K_i/S_i)	6/2	2/0	15/3	15/2	14/14	-	-26.15

gorithm appears to have "repaired" the infeasible solution found by the standard genetic algorithm. The average percentage of infeasible solutions in the final generations of the genetic algorithm runs was equal to 7.2%. Table 3 contains the objective function values $E\left[f(\mathbf{x})\right]^*$ and service levels $E\left[SL(\mathbf{x})\right]^*$ % with 95% confidence bounds of the best parameters found by both the standard genetic algorithm and the hybrid genetic algorithm with the "death" penalty function.

These data were produced by running 50 replicates of each of the six simulation models for $t_{sim} = 1,500,000.0$ time units and then averaging the corresponding variables. This is a 18.75 times

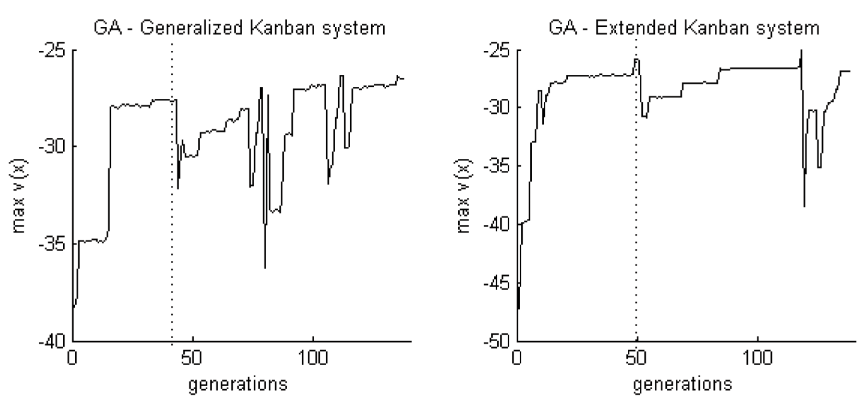
Table 3. Objective function values $E\left[f(\mathbf{x})\right]^*$ and % service levels $E\left[SL(\mathbf{x})\right]^*$ % for best parameter sets found by standard GA and hybrid GA with "death penalty" (95% confidence). (K stands for Kanban, BS for Base Stock etc. n.i. stands for "no improvement")

GA results			**Hybrid GA (with local search)**		
Policies	$E\left[SL(\mathbf{x})\right]^*$ %	$E\left[f(\mathbf{x})\right]^*$	**Policies**	$E\left[SL(\mathbf{x})\right]^*$ %	$E\left[f(\mathbf{x})\right]^*$
K	91.17 ± 0.09	-32.07 ± 0.04	**K**	90.08 ± 0.09	-29.13 ± 0.03
BS	90.37 ± 0.09	-31.73 ± 0.02	**BS**	90.03 ± 0.11	-29.53 ± 0.02
C	89.87 ± 0.08	-25.26 ± 0.01	**C**	n.i.	n.i.
C/K H	91.34 ± 0.08	-28.83 ± 0.03	**C/K H**	90.43 ± 0.11	-26.79 ± 0.04
EK	90.29 ± 0.07	-26.57 ± 0.0	**EK**	90.21 ± 0.09	-26.23 ± 0.03
GK	90.23 ± 0.09	-28.18 ± 0.02	**GK**	89.91 ± 0.01	-26.12 ± 0.02

longer simulation than that used to evaluate the fitness of the individuals in the genetic algorithm. By using exhaustively long simulation times we can compute far more accurate estimators, (indicated by the superscript *), which can be considered to approximate the true expected values of these performance measures. This way, relatively safe conclusions can be drawn regarding both the quality of the solutions found by the hybrid genetic algorithms and the performance of each of the competing pull type control policies. Of course, by increasing the simulation time and/or the resampling, the optimization algorithm is less likely to be mislead by "lucky" candidate solutions which score well once by chance but the consequent computational cost is prohibitive of doing so. Apart from that, we are interested in establishing whether the algorithm is capable of locating good and hopefully optimal solutions in the presence of a relatively low signal-to-"noise" ratio. By observing the data in Table 3 we see that the local search algorithm produced actual improvements in the objective function values while preserving the feasibility of the solutions in all cases except the CONWIP and Generalized Kanban systems. The results regarding the Generalized Kanban system are somehow contradicting. In Table 2 the local search algorithm appears to have repaired the infeasible solution, while in Table 3 the original solution is now found to be actually feasible and the local search solution is the one which violates the constraint. Actually, these results merely demonstrate an inherent weakness of search algorithms that generate a single point per iteration when compared to genetic algorithms in "noisy" environments. A hill-climbing method, like the one used here, compares candidate solutions in each iteration only with best solution found so far, and therefore, it is easy to be mislead by a "lucky" solution. In genetic algorithms, on the contrary, for a solution to be maintained it must outweigh an entire collection of solutions and not just a previously best one. As an overall assessment, we could argue that the results presented in this section support the conclusion that even with this simple static penalty function a genetic algorithm can produce quite good solutions. Two typical plots of the best solution found by the genetic algorithm with the "death" penalty function versus the number of generations can be found in Figure 2.

These two plots exhibit a somehow similar pattern. For illustration purposes we use a time window of 140 generations and divide the plot areas in two regions with a perpendicular dotted line. Notice that in the left side of the plots, if we disregard random fluctuations caused by the simulation model, the two curves are increasing

Figure 2. Typical plots of best fitness value found by GA with "death" penalty function versus number of iterations (140 iterations window)



almost monotonically. In this area, the best solution found by the algorithm is not lying somewhere near the boundaries between feasible and infeasible region. In the intersection point of the curve with the dotted perpendicular line the curve suddenly "dives". This is indicative that the currently best individual is marginally feasible, (or infeasible), and that it failed to satisfy the constraint in this evaluation. As a consequence, it was penalized heavily and substituted by another individual which happened to have a lower fitness value. From this point on, the curve displays similar abrupt fluctuations, indicating that the population evolves towards the boundaries between feasible and infeasible region and the optimal solution.

DESIGNING EXPONENTIAL PENALTY FUNCTION

Intuitively, the "death" penalty approach, as attractive as it can be due to its simplicity, does not seem to be the best approach to handle constraints. Even the slightest violation of the imposed constraints results in penalizing heavily a good solution. This way an individual which scores excellently for a series of consecutive generations may be discarded by the algorithm. This is an undesired property in the kind of optimization problem that we are dealing with, where fitness measurements are distorted by random fluctuations caused by the stochasticity of the simulation model. Another weakness of this approach is that it damages the diversity of the population, as the majority of the individuals are crowded in the feasible region. Given that the optimal solution lies on the feasibility boundaries, the search would probably be more efficient if the population evolved towards the boundaries from both feasible and infeasible regions. For example, it is unclear why a slightly infeasible individual which is located very close to the optimal solution should be assigned a worse fitness value than a feasible individual that scores poorly. For all of the above mentioned reasons, the idea of penalizing infeasible solutions according to the level of the constraint violation seems more appealing, (see Venkatraman and Yen (2005) for guidelines on designing penalty functions). The problem that needs to be addressed now is how to design such a "soft-limit" penalty function. A reasonable choice is to use an exponential penalty function $p(\mathbf{x}) = c^u$, where $c = const \in \Re$ and $u = t - SL$ is the difference between the target service level and $E\left[SL(\mathbf{x})\right]$ the measured expected service level. The intuitive, *minimal penalty rule*, (Le Riche *et al.* 1995), suggests that the penalty for infeasible individuals should be just above the threshold below which infeasible solutions score better than their feasible, possibly optimal, neighbors. In practice, however, it is quite difficult to achieve this. The procedure we followed in order to implement, at least to some extent, this intuition, is the following. Using the output of the executions of the genetic algorithm with the "death" penalty function we created plots like the ones in Figure 2. By examining these plots it was easy to locate solutions that were very close to the feasibility boundaries, (these points are indicated by the characteristic "dive" of the curve). The next step was to examine the neighborhood of such a point in order to determine how a small change in parameters affected the service level *SL* as well as the objective function $f(\mathbf{x})$. The value of *SL* is affected primarily by the control parameters of the last three machines so we could limit ourselves to a relatively small neighborhood. Having collected this data, we were able to select the parameter c of the penalty function $p(\mathbf{x}) = c^u$, in the spirit of the "minimal penalty rule". This is an empirical technique that may not be easy or even possible to apply to other problems, nevertheless it provides the means to design a penalty function that will work well and outperform most of the times the "death" penalty approach as supported by our experimental results presented in the following section.

EXPONENTIAL PENALTY FUNCTION RESULTS

After following the procedure outlined in the previous section we were able to construct the following penalty function (6).

$$p(\mathbf{x}) = \begin{cases} 0.0, & u \leq 0.0 \\ 9.0^{u}, & 0.0 < u < 3.0 \\ 9.0^{3}, & 3.0 \leq u \end{cases} \quad (6)$$

where $u = t - E\left[SL(\mathbf{x})\right]$ is the difference between the target service level t=90.0% and the measured service level $E\left[SL(\mathbf{x})\right]$. Note that for service levels equal to or lower than 87.0% we ground the penalty to a constant value. The reason we do this is that we want the raw fitness values $v(\mathbf{x})$ to be within a range for the selection operator of the genetic algorithm to work properly. We reiterate that in our implementation the values of the individuals are scaled proportionally to their raw fitness measurements prior to selection.

The results from the hybrid genetic algorithm runs for each control policy are displayed in Table 4. The rows containing the results of the standard genetic algorithm, (without the local search component), are labeled with the initials GA followed by the control policy's name, while the results of the hybrid algorithm are labeled with the initials GAL and the control policy's name. The last column of Table 4 contains the fitness values of the corresponding parameter sets. The CONWIP system scored the highest fitness value $v_b = -25.31$ followed by the Extended Kanban, the Hybrid CONWIP/Kanban, the Generalized Kanban, the Base Stock and the Kanban systems in decreasing fitness value order. The hybrid optimization algorithm outperformed the standard genetic algorithm in the cases of the Base Stock, the Extended Kanban and the Generalized Kanban systems. For the three remaining systems the local search failed to offer an improvement in minimizing $v\left(\mathbf{x}\right)$. Table 5 contains the objective function values $E\left[f(\mathbf{x})\right]^*$ and service levels $E\left[SL(\mathbf{x})\right]^*$ % with 95% confidence bounds of the best parameters found by both

Table 4. Best parameter sets and fitness values for exponential penalty function (n.i. stands for "no improvement")

Policies	*$x_1(/x_1')$*	*$x_2(/x_2')$*	*$x_3(/x_3')$*	*$x_4(/x_4')$*	*$x_5(/x_5')$*	*C*	*v(x)*
GA_Kanban (Ki)	1	1	1	7	13	-	-28.05
GAL_Kanban (Ki)	n.i.	n.i.	n.i.	n.i.	n.i.	-	n.i.
GA_BaseStock (Si)	0	3	0	0	17	-	-27.19
GAL_BaseStock (Si)	0	2	0	0	17	-	-25.95
GA_CONWIP (Si, C)	0	6	1	6	6	19	-25.31
GAL_CONWIP (Si, C)	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
GA_Hybrid (Ki, i=1,2,3,4, B5, C)	1	4	5	9	1	20	-25.92
GAL_Hybrid(Ki,i=1,2,3,4, S5, C)	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
GA_E. Kanban (Ki/Si)	4/1	10/0	2/2	4/2	22/15	-	-25.82
GAL_E. Kanban (Ki/Si)	3 /1	10/0	2/2	4/2	22/15	-	-25.72
GA_G. Kanban (Ki/Si)	12/5	7/0	14/4	13/1	16/14	-	-29.34
GAL_G.Kanban (Ki/Si)	9/2	3/0	13/4	13/1	15/14	-	-25.95

Table 5. Objective function values $E\left[f(\mathbf{x})\right]^*$ and % service levels $E\left[SL(\mathbf{x})\right]^*$ % for best parameter sets found by standard GA and hybrid GA with "exponential penalty" (95% confidence). (K stands for Kanban, BS for Base Stock etc. n.i. stands for "no improvement")

GA results			Hybrid GA (with local search)		
Policies	$E\left[SL(\mathbf{x})\right]^*$ %	$E\left[f(\mathbf{x})\right]^*$	**Policies**	$E\left[SL(\mathbf{x})\right]^*$ %	$E\left[f(\mathbf{x})\right]^*$
K	90.10 ± 0.08	-28.01 ± 0.03	**K**	n.i.	n.i.
BS	90.34 ± 0.12	-27.20 ± 0.03	**BS**	89.95 ± 0.13	-25.97 ± 0.03
C	89.94 ± 0.10	-25.27 ± 0.02	**C**	n.i.	n.i.
C/K H	90.21 ± 0.09	-25.87 ± 0.02	**C/K H**	n.i.	n.i.
EK	90.07 ± 0.11	-25.84 ± 0.02	**EK**	89.98 ± 0.10	-25.71 ± 0.02
GK	90.31 ± 0.09	-29.34 ± 0.02	**GK**	89.73 ± 0.09	-25.89 ± 0.02

the standard genetic algorithm and the hybrid genetic algorithm with the exponential penalty function. These data were produced by running 50 replicates of each of the six simulation models for $t_{sim} = 1,500,000.0$ time units and then averaging the corresponding variables.

All three solutions found by the local search algorithm when initialized with the solutions of the genetic algorithm were marginally infeasible. The local search algorithm falsely interpreted the effect of random noise as an actual improvement and thus substituted the feasible solutions by infeasible ones. Of course, we cannot rule out that this was caused in part by the penalty function itself. However, we must mention that the amount of the constraint violation was rather trivial. The average percentage of infeasible solutions in the final generations of the genetic algorithm runs with the exponential penalty function was equal to 8.5%.

DISCUSSION ON THE PERFORMANCE OF THE TWO PENALTY FUNCTIONS

By comparing the data in Tables 3 and 5 we notice that the standard genetic algorithm with the exponential penalty function outperforms both the standard genetic algorithm and the hybrid algorithm with the "death penalty" function for all systems except the CONWIP and the Generalized Kanban. In terms of objective function value, the use of the exponential penalty function rather than the "death penalty" improved the solution by 3.84% for the Kanban system, by 7.89% for the Base Stock system and by 3.43% for the CONWIP/Kanban Hybrid system. For the Extended Kanban system we monitored a 1.5% lower value of $E\left[f(\mathbf{x})\right]^*$, while for the CONWIP system the results were practically the same. Only for the Generalized Kanban system the "death" penalty approach succeeded in producing a 4% better solution than the exponential penalty approach. The superiority of the exponential penalty function over the "death" penalty function can be explained qualitatively as follows: Figure 3 shows typical plots of the genetic algorithm's convergence with the "death" penalty and the exponential penalty function. Notice that at some point near the 60th generation both curves have approximately the same height. The best solutions found by the two implementations of the algorithm in these points probably belong to the same level set and are lying somewhere close to the feasibility boundaries. In some subsequent iteration of the algorithm with the "death" penalty, this solution apparently violates the constraint and is therefore discarded. The height of the curve shows that the individual which replaced it has a significantly lower fitness value. This is not the case in the algorithm with the exponential penalty where the properly designed penalty function prevents the good solution to be discarded, at least not from a much worse candidate solution. Concluding the discussion on the performance of the hybrid GA we state summarize our major findings: i) the incorporation of a local search element can enhance the genetic algorithm's performance with the disadvantage of that the local search algorithm is more susceptible to falsely interpreting random noise as actual objective function improvements than the genetic algorithm, ii) the "death penalty" approach most of the times will yield worse results than a function which penalizes solutions according to the level of the constraint violation like the exponential penalty function used here.

COMPARISON OF PULL TYPE PRODUCTION CONTROL POLICIES-SENSITIVITY ANALYSIS

Table 6 presents the objective function values and the corresponding service levels for the six JIT control policies with the best parameters found by the proposed optimization strategy. Note that the Base stock, CONWIP and Extended Kanban solutions attain a service level below 90% but since the constraint is within the 95% confidence halfwidth we consider them to be feasible. The CONWIP policy ranks first followed in close distance by the Extended Kanban, Hybrid and Base Stock policies. The Kanban and Generalized Kanban policies occupy the last two positions of the objective function value ranking. Since in this simulation scenario the demand process pushes the manufacturing system towards its maximum throughput rate, the poor performance of the Kanban mechanism is anticipated since this policy offers tight coordination between the manufacturing stages but does not respond rapidly to incoming orders. On the other hand the performance of the Generalized Kanban system is somewhat unexpected since it is supposed to be an enhancement of the original Kanban policy. However, this is not the case for the control policy that is mostly related to Generalized Kanban, the Extended Kanban mechanism, which ranks second. The main characteristic of the Base Stock policy, that is fast reaction to demand, is supported by the experimental output. Finally, the fact that in a CONWIP or CONWIP/Kanban Hybrid system the WIP tends to accumulate to the last buffer al-

Figure 3. Typical plots of best fitness value found by GA with "death" and exponential penalty functions versus number of iterations

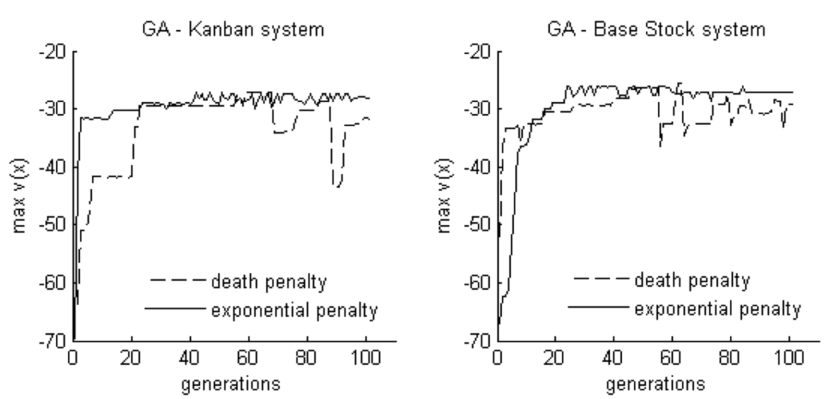


Table 6. Objective function values – service levels of pull control policies with best parameters for base simulation case

	Kanban	Base Stock	CONWIP	CONWIP/Kanban Hybrid	Extended Kanban	Generalized Kanban
$E\left[f(\mathbf{x})\right]^*$	-28.01±0.03	-25.97±0.03	-25.27±0.02	-25.87±0.02	-25.71±0.02	-28.18±0.02
$E\left[SL(\mathbf{x})\right]^*$	90.10 ± 0.08	89.95 ± 0.13	89.94 ± 0.10	90.21 ± 0.09	89.98 ± 0.10	90.23 ± 0.09

lows this two policies to achieve a high service level while operating a lean manufacturing mode.

Table 7 shows the statistics of the system's performance measures for the four variants of the basic simulation case. In the case where the demand rate increases (first column of Table 7) we notice that the service level as well as the average WIP decreases for all policies, but some control mechanisms are more sensitive to this change than others. Specifically, the service rate in the Kanban and Hybrid systems decreases dramatically, whereas the Base Stock and Generalized Kanban policies seem to be more robust regarding the increase of the demand rate. In the second variant (decreased arrival rate) of the basic simulation case one can see that all six control policies practically achieve the same service level. This is an indication that when the demand can be easily satisfied by the manufacturing system the role of the production control policy diminishes. In this case the distribution of the objective function values over the control mechanisms also tends to level out.

The increase of the standard deviation of the processing times (variant 3) has an effect similar to that of the increase of the demand rate. The reasons for that can be attributed to the resulting decreased coordination among the various production stages which increases the frequency with which machine starvation or blockage events occur. Again, the Kanban mechanism is mostly affected by this parameter due to its tight production coordination scheme, while the Generalized Kanban mechanism seems to react rather ro-

Table 7. Objective function values – service levels of pull control policies with best parameters for variants of base simulation case

	R_a=1.0		***R_a=0.8***		***st.d.=0.1***		***st.d.=0.001***	
	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$
K	-15.21	55.61	-32.77	96.52	-20.92	76.2	-28.32	90.43
BS	-23.75	64.2	-28.57	96.05	-25.48	88	-26.01	90.07
C	-19.15	62.89	-28.68	96.18	-24.54	87.91	-25.33	90.12
H	-14.71	57.91	-30.06	96.38	-21.23	79.43	-26.03	90.42
EK	-18.05	61.89	-29.33	96.21	-24.96	88.05	-25.74	90.07
GK	-20.57	62.92	-31.74	96.29	-27.56	88.72	-28.17	90.27

bustly. Finally, the decrease of the standard deviation of the processing times (variant 4) seems to have a negligible effect on the system's behavior as indicated by the experimental data presented in the last column of Table 7. Tables 8 and 9 contain data regarding the sensitivity of the system's behavior in respect to the parameters of the controlling policy. For example, in Table 8, the cells in the *i*-th row that belong to the columns labeled as "Base Stock" show the objective function value and service level that result when the *i*-th component (parameter S_i) of the corresponding decision variable vector is increased by the minimum possible value, i.e. by one. In general, the service level (objective function) is an increasing (decreasing) function of the control parameters. However, the rate with which the service level/objective function changes depends on the type of the control policy and the index (position) of the parameter in the parameter vector. For instance, in the five-station Kanban system, adding an additional kanban in the last stage will result in a larger decrease of the objective function value than adding an extra kanban in any of the upstream stages. It is interesting to observe the cases of the CONWIP and CONWIP/Kanban Hybrid systems where the unitary increase of a control parameter in any of the stages 2,3,4,5 seems to have the same effect. This can be explained by the fact that since the last workstation is authorized to produce whenever it has this ability all parts in upstream buffers are continuously "pushed" towards the finished goods buffer, and therefore WIP in intermediate stages is scarce. By increasing the initial stock in the first buffer the average WIP in intermediate stages increases and thus this change has greater impact to the objective value and service level. Generalized Kanban and Extended Kanban are characterized by two parameters per stage and therefore the sensitivity analysis must consider both of these parameters. The systems' performance for a unitary change in the base stock of the *i*-th stage is shown in the columns labeled as "base stocks" whereas the cells under the label "free kanbans" contain system performance information when the total number of kanbans of the *i*-th stage is increased by one but the base stock remains unaltered.

The effect of adding an extra base stock to a stage of a Generalized/Extended Kanban system is similar to that of adding a kanban to a stage of a Kanban system. The mean WIP is also an increasing function of the control parameters (K_i - S_i) but as one can see from Table 9 rather large changes in the number of free kanbans are needed for a significant change in the objective function value to occur.

Table 8. Objective function value – service level sensitivity to parameter vector (Kanban, Base Stock, CONWIP, Hybrid)

	Kanban		**Base Stock**		**CONWIP**		**CONWIP/Kanban Hybrid**	
	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$
1	-29.1	90.32	-27.15	90.35	-27.53	90.92	-28.03	91.13
2	-29.34	90.51	-27.19	90.26	-27.24	90.82	-27.84	91.02
3	-29.52	90.65	-27.77	90.80	-27.24	90.77	-27.83	91.02
4	-29.61	90.51	-27.83	90.74	-27.24	90.81	-27.79	91.0
5	-29.86	90.94	-27.86	90.78	-27.24	90.83	-27.76	90.99

Table 9. Objective function value – service level sensitivity to parameter vector (Extended Kanban, Generalized Kanban)

	Extended Kanban				Generalized Kanban			
	base stocks (S_i)		free kanbans (K_i-S_i)		base stocks (S_i)		free kanbans (K_i-S_i)	
	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$	$E\left[f(\mathbf{x})\right]^*$	$E\left[SL(\mathbf{x})\right]^*$
1	-26.75	90.21	-25.83	90.07	-29.18	90.35	-28.20	90.27
2	-27.12	90.47	-25.79	90.01	-29.65	90.63	-28.27	90.29
3	-27.16	90.55	-25.8	90.1	-29.58	90.69	-28.29	90.24
4	-27.38	90.66	-25.77	90.06	-29.81	90.78	-28.26	90.28
5	-27.61	90.86	-25.73	90.06	-30.08	91.13	-28.21	90.32

CONCLUSION AND FUTURE RESEARCH

We implemented a hybrid optimization technique which combines a genetic algorithm with a local search procedure to find optimal decision variables for a family of JIT manufacturing systems. The goal was to maximize a weighted sum of the mean WorkInProcess inventories subject to the constraint of maintaining a target service level. Our numerical results indicate that the performance of a genetic algorithm can be easily enhanced by incorporating a local search component, however, the local search algorithm is more susceptible to falsely interpreting random noise as actual objective function improvements than the genetic algorithm. Moreover, our results support the intuitive perception that penalizing candidate solutions according to the level of constraint violation will yield better results than the "death penalty" approach most of the times. The performance of the JIT control policies with optimized parameters is presented analytically and commented upon. Finally, we conduct a sensitivity analysis in respect to the variation of the demand rate, the standard deviation of the service rates and the control parameter vector. The results of the analysis offer considerable insight to the underlying mechanics of the JIT control policies under consideration. Constraint handling and "noisy" or dynamic environments in the context of genetic optimization of manufacturing systems are currently active research fields. Indicatively, a relatively recent and interesting direction is to use evolutionary multi-objective techniques to handle constraints as additional objectives.

REFERENCES

Alabas, C., Altiparmak, F., & Dengiz, B. (2002). A comparison of the performance of artificial intelligence techniques for optimizing the number of kanbans. *The Journal of the Operational Research Society*, *53*(8), 907–914. doi:10.1057/palgrave.jors.2601395

Bellman, R. E. (1957). *Dynamic Programming*. Princeton: Princeton University Press.

Berkley, B. J. (1992). A review of the kanban production control research literature. *Production and Operations Management*, *1*(4), 393–411. doi:10.1111/j.1937-5956.1992.tb00004.x

Bhat, U. N. (2008). *An Introduction to Queueing Theory: Modelling and Analysis in Applications*. Boston: Birkhauser.

Bowden, R. O., Hall, J. D., & Usher, J. M. (1996). Integration of evolutionary programming and simulation to optimize a pull production system. *Computers & Industrial Engineering*, *31*(1&2), 217–220. doi:10.1016/0360-8352(96)00115-5

Buzacott, J. A., & Shanthikumar, J. G. (1992). A general approach for coordinating production in multiple cell manufacturing systems. *Production and Operations Management*, *1*(1), 34–52. doi:10.1111/j.1937-5956.1992.tb00338.x

Buzacott, J. A., & Shanthikumar, J. G. (1993). *Stochastic Models of Manufacturing Systems*. New York: Prentice Hall.

Dallery, Y., & Liberopoulos, G. (2000). Extended kanban control system: combining kanban and base stock. *IIE Transactions*, *32*(4), 369–386. doi:10.1080/07408170008963914

Fitzpatrick, J. M. & Grefenstette, J. J. (1988) Genetic algorithms in noisy environments. *Machine Learning: Special issue on genetic algorithms, 3*, 101-120.

Gershwin, S. B. (1994). *Manufacturing Systems Engineering*. New York: Prentice Hall.

Hammel, U., & Bäck, T. (1994). Evolution strategies on noisy functions. How to improve convergence properties. *Parallel Problem Solving from Nature*, *3*, 159–168.

Howard, R. (1960). *Dynamic Programming and Markov Processes. Massachussets*. MIT Press.

Hurrion, R. D. (1997). An example of simulation optimization using a neural network metamodel: finding the optimum number of kanbans in a manufacturing system. *The Journal of the Operational Research Society*, *48*(11), 1105–1112.

Karaesmen, F., & Dallery, Y. (2000). A performance comparison of pull type control mechanisms for multi-stage manufacturing. *International Journal of Production Economics*, *68*, 59–71. doi:10.1016/S0925-5273(98)00246-1

Koulouriotis, D. E., Xanthopoulos, A. S., & Gasteratos, A. (2008). A Reinforcement Learning Approach for Production Control in Manufacturing Systems. In *1st International Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems*, (pp. 24-31). Patra, Greece.

Koulouriotis, D. E., Xanthopoulos, A. S., & Tourassis, V. D. (2010). Simulation optimisation of pull control policies for serial manufacturing lines and assembly manufacturing systems using genetic algorithms. *International Journal of Production Research*, *48*(10), 2887–2912. doi:10.1080/00207540802603759

Law, A., & Kelton, D. (2000). *Simulation Modelling and Analysis*. New York: McGraw Hill.

Le Riche, R., Knopf-Lenoir, C., & Haftka, R. T. (1995). A Segregated Genetic Algorithm for Constrained Structural Optimization. In *Sixth International Conference on Genetic Algorithms* (pp. 558-565).

Liberopoulos, G., & Dallery, Y. (2000). A unified framework for pull control mechanisms in multi-stage manufacturing systems. *Annals of Operations Research*, *93*, 325–355. doi:10.1023/A:1018980024795

Panayiotou, C. G., & Cassandras, C. G. (1999). Optimization of kanban - based manufacturing systems. *Automatica*, *35*(3), 1521–1533. doi:10.1016/S0005-1098(99)00074-6

Paternina – Arboleda, C. D, & Das, T. K. (2001). Intelligent dynamic control policies for serial production lines. *IIE Transactions*, *33*, 65–77. doi:10.1080/07408170108936807

Shahabudeen, P., Gopinath, R., & Krishnaiah, K. (2002). Design of bi-criteria kanban system using simulated annealing technique. *Computers & Industrial Engineering*, *41*(4), 355–370. doi:10.1016/S0360-8352(01)00060-2

Shahabudeen, P., & Krishnaiah, K. (1999). Design of a Bi-Criteria kanban system using Genetic Algorithm. *International Journal of Management and System*, *15*(3), 257–274.

Smith, G. C., & Smith, S. S. F. (2002). An enhanced genetic algorithm for automated assembly planning. *Robotics and Computer-integrated Manufacturing*, *18*(5-6), 355–364. doi:10.1016/S0736-5845(02)00029-7

Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). CONWIP: a pull alternative to kanban. *International Journal of Production Research*, *28*, 879–894. doi:10.1080/00207549008942761

Sugimori, Y., Kusunoki, K., Cho, F., & Uchikawa, S. (1977). Toyota production system and kanban system materialization of just-in-time and respect-for-humans systems. *International Journal of Production Research*, *15*(6), 553–564. doi:10.1080/00207547708943149

Veatch, M. H., & Wein, L. M. (1992). Monotone control of queueing networks. *Queueing Systems*, *12*, 391–408. doi:10.1007/BF01158810

Venkatraman, S., & Yen, G. G. (2005). A generic framework for constrained optimization using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, *9*(4), 424–435. doi:10.1109/TEVC.2005.846817

Vivo-Truyols, G., Torres-Lapasio, J. R., & Garcıa-Alvarez-Coque, M. C. (2001). A hybrid genetic algorithm with local search: I. Discrete variables: optimisation of complementary mobile phases. *Chemometrics and Intelligent Laboratory Systems*, *59*, 89–106. doi:10.1016/S0169-7439(01)00148-4

Yamamoto, H., Qudeiri, J. A., & Marui, E. (2008). Definition of FTL with bypass lines and its simulator for buffer size decision. *International Journal of Production Economics*, *112*(1), 18–25. doi:10.1016/j.ijpe.2007.03.007

Yang, T., Kuo, Y., & Cho, C. (2007). A genetic algorithms simulation approach for the multi-attribute combinatorial dispatching decision problem. *European Journal of Operational Research*, *176*(3), 1859–1873. doi:10.1016/j.ejor.2005.10.048

Yuan, Q., He, Z., & Leng, H. (2008). A hybrid genetic algorithm for a class of global optimization problems with box constraints. *Applied Mathematics and Computation*, *197*, 924–929. doi:10.1016/j.amc.2007.08.081

This work was previously published in Supply Chain Optimization, Design, and Management: Advances and Intelligent Methods, edited by Ioannis Minis, Vasileios Zeimpekis, Georgios Dounias and Nicholas Ampazis, pp. 212-231, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 38
Comparison of Connected vs. Disconnected Cellular Systems:
A Case Study

Gürsel A. Süer
Ohio University, USA

Royston Lobo
S.S. White Technologies Inc., USA

ABSTRACT

In this chapter, two cellular manufacturing systems, namely connected cells and disconnected cells, have been studied, and their performance was compared with respect to average flowtime and work-in-process inventory under make-to-order demand strategy. The study was performed in a medical device manufacturing company considering their a) existing system b) variations from the existing system by considering different process routings. Simulation models for each of the systems and each of the options were developed in ARENA 7.0 simulation software. The data used to model each of these systems were obtained from the company based on a period of nineteen months. Considering the existing system, no dominance was established between connected cells vs. disconnected cells as mixed results were obtained for different families. On the other hand, when different process routings were used, connected system outperformed the disconnected system. It is suspected that one additional operation required in the disconnected system as well batching requirement at the end of packaging led to poor performance for the disconnected cells. Finally, increased routing flexibility improved the performance of the connected cells, whereas it had adverse effects in the disconnected cells configuration.

DOI: 10.4018/978-1-4666-1945-6.ch038

INTRODUCTION

Cellular Manufacturing is a well known application of Group Technology (GT). Cellular Design typically involves determining appropriate part families and corresponding manufacturing cells. This can be done either by grouping parts into families and then forming machine cells based on the part families or machine cells are determined first and based on these machine cells the part families may be formed or lastly both these formations can take place simultaneously. In a cellular manufacturing system, there may be a manufacturing cell for each part family or some of the manufacturing cells can process more than one part family based on the flexibility of the cells. The factors affecting the formation of cells can differ under various circumstances, some of them are volume of work to be performed by the machine cell, variations in routing sequences of the part families, processing times, etc.

A manufacturing system in which the goods or products are manufactured only after customer orders are received is called a make-to-order system. This type of system helps reduce inventory levels since no finished goods inventory is kept on hand.

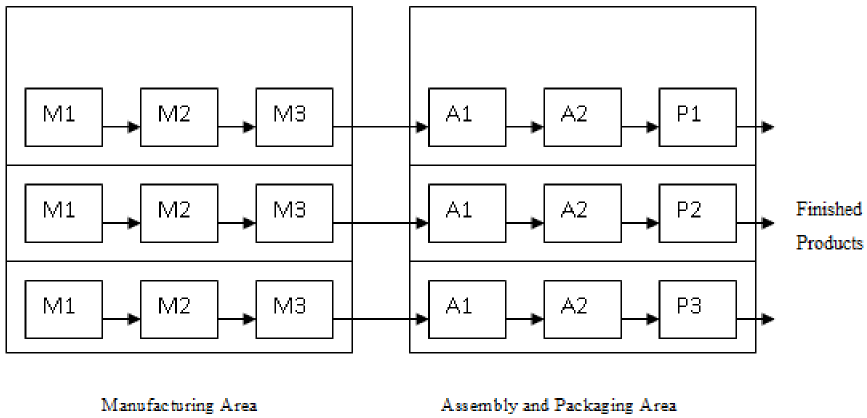
In this chapter, two types of cellular layouts are analyzed, namely connected cells (single-stage cellular system) and disconnected cells (multi-stage cellular system) and their performance is compared under various circumstances for a make-to-order company. This problem has been observed in a medical device manufacturing company. The management was interested in such a comparison to finalize the cellular design. It was also important to research the impact of flexibility within each system for different combinations of family routings. A similar situation of connected vs. disconnected cellular design was also observed in a shoe manufacturing company, and in a jewelry manufacturing company. Authors believe that this problem has not been addressed in the literature before even though it has been observed in more than one company and therefore worthy to study.

BACKGROUND

The connected cells represent a continuous flow where the products enter the cells in the manufacturing area, complete the machining operations and exit through the corresponding assembly and packaging area after completion of the assembly and packaging operations. In other words, the output of a cell in the manufacturing area becomes the input to the corresponding cell in the assembly and packaging area. The biggest advantage of connected cells is that material flow is smoother and hence flowtime is expected to be shorter. This is also expected to result in lower WIP inventory. This paper focuses on a cellular manufacturing system similar to the system shown in Figure 1. There are three cells in the manufacturing area and three cells in the assembly and packaging area. In these cells, M1 through M3 represent the machines in the manufacturing area, A1, A2 and P1 through P3 represent the machines in the assembly and packaging area. The products essentially follow a unidirectional flow. The three cells in manufacturing area are similar since they have similar machines and all the products can be manufactured in any of the cells. However, the situation gets complicated in the assembly and packaging area. The three cells have restrictions in terms of the products that they can process. Therefore, deciding which manufacturing cell a product should be assigned is dictated by the packaging cell(s) it can be processed later on. This constraint makes the manufacturing system less flexible.

In the disconnected cell layout, the products enter the manufacturing area, complete the machining operations and exit this area. On exiting the manufacturing area, the products can go to more than one of the assembly and packaging cells. In other words, the output from the cells in

Figure 1. Connected cells



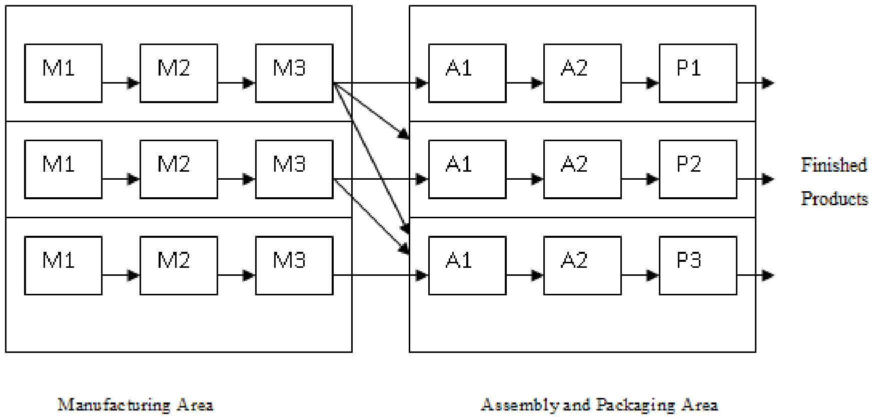
the manufacturing area can become an input for some of the cells in the assembly and packaging area (partially flexible disconnected cells) or all of them (completely flexible disconnected cells). Figure 2 shows a partially flexible disconnected cells case where the parts from cell 1 in the manufacturing area can go to any of the cells in the assembly and packaging area. Parts from cell 2 can only go to cell 2, and cell 3 of the assembly and packaging area. Parts from cell 3 of the manufacturing area can only go to cell 3 of the assembly and packaging area. The disconnected system design allows more flexibility. On the other hand, due to interruptions in the flow, some delays may occur which may eventually lead to higher flowtimes and WIP inventory levels.

LITERATURE REVIEW

A group of researchers compared the performance of cellular layout with process layout. Flynn and Jacobs (1987) developed a simulation model using SLAM for an actual shop to compare the performance of group technology layout against process layout. Morris and Tersine (1990) developed simulation models for a process layout and a cellular layout using SIMAN. The two performance measures used were throughput time and work-in-process inventory (WIP). Yazici (2005) developed a simulation model using Promodel based on data collected from a screen-printing company to ascertain the influence of volume, product mix, routing and labor flexibilities in the presence of fluctuating demand. A comparison between one-cell, two-cell configurations versus a job shop is made to determine the shortest delivery and highest utilization. Agarwal and Sarkis (1998) reviewed the conflicting results from the literature in regard to superiority of cellular layout vs. functional layout. They attempted to identify and compile the existing studies and understand conflicting findings. Johnson and Wemmerlov (1996) analyzed twenty-four model-based studies and concluded that the results of these work cannot assist practitioners in making choices between existing layouts and alternative cell systems. Shafer and Charnes (1993) studied cellular manufacturing under a variety of operating conditions. Queueing theoretic and simulation models of cellular and functional layouts are developed for various shop operating environments to investigate several factors believed to influence the benefits associated with a cellular manufacturing layout.

Another group of researchers focused on analyzing cellular systems. Selen and Ashayeri (2001) used a simulation approach to identify improvements in the average daily output through

Figure 2. Disconnected cells with partial flexibility



management of buffer sizes, reduced repair time, and cycle time in an automotive company. Albino and Garavelli (1998) simulated a cellular manufacturing system using Matlab to study the effects of resource dependability and routing flexibilities on the performance of the system. Based on the simulation results, the authors concluded that as resource dependability decreases, flexible routings for part families can increase productivity. On the contrary, from an economic standpoint they concluded that benefits will greatly reduce from an increase routing flexibility cost and resource dependability. Caprihan and Wadhwa (1997) studied the impact of fluctuating levels of routing flexibility on the performance of a Flexible Manufacturing System (FMS). Based on results obtained, the authors concluded that there is an optimal flexibility level beyond which the system performance tends to decline. Also, increase in routing flexibility when made available with an associated cost seldom tends to be beneficial. Suer, Huang, and Maddisetty (2009) discussed layered cellular design to deal with demand variability. They proposed a methodology to design a cellular system that consisted of dedicated cells, shared cells and remainder cell.

Other researchers studied make-to-order and make-to-stock production strategies. Among them, DeCroix and Arreola-Risa (1998) studied the optimality of a Make-to- Order (MTO) versus a Make-to-Stock (MTS) policy for a manufacturing set up producing various heterogeneous products facing random demands. Federgruen and Katalan (1999) investigated a hybrid system comprising of a MTO and a MTS systems and presented a host of alternatives to prioritize the production of the MTO and MTS items. Van Donk (2000) used the concept of decoupling point (DP) to develop a frame in order to help managers in the food processing industries to decide which of their products should be MTO and which ones should be MTS. Gupta and Benjaafar (2004) presented a hybrid strategy which is a combination of MTO and MTS modes of production. Nandi and Rogers (2003) simulated a manufacturing system to study its behavior in a make to order environment under a control policy involving an order release component and an order acceptance/rejection component.

Authors are not aware of any other study that focuses on comparing the performance of connected cells with disconnected cells and therefore we believe this is an important contribution to the literature.

DESCRIPTION OF THE SYSTEM STUDIED: THE CASE STUDY

This section describes the medical device manufacturing company where the experimentation was carried out. The products essentially follow a unidirectional flow. The manufacturing process is mainly divided into two areas, namely fabrication and packaging. Each area consists of three cells and cells are not identical. The one piece-flow strategy is adapted in all cells. The company has well defined families which are determined based on packaging requirements. Furthermore, the cells have been already formed. The average flowtime and the work-in-process inventory are the performance measures used to evaluate the performance of connected cells and disconnected cells.

Product Families

The products are grouped under three families: Family 1 (F1), Family 2 (F2), and Family 3 (F3). The finished products are vials consisting of blood sugar strips and each vial essentially contains 25 strips. The number of products in families 1, 2 and 3 are 11, 21 and 4, respectively.

The families that are described were already formed by the manufacturer based on the number of vials (subfamilies) included in the box. Family 1 requires only one subassembly (S), one box (B1), one label (L), and one Insert for instructions (I); family 2 (F2) requires 2 subassemblies, one box (B2), one label and one insert, and family 3 (F3) requires 4 subassemblies, one box (B3), one label and one insert to become finished product as shown in Table 1. Obviously, this family classification is strictly from manufacturing perspective and marketing department uses its own family definition based on product function related characteristics. The family definition has been made based on limitations of packaging machines. Not all packaging machines can insert 4 vials into a box. This seemingly simple issue becomes an obstacle in assigning products to packaging cells and furthermore becomes a restriction in assigning products to even manufacturing cells in connected cellular design.

Fabrication Cells

The fabrication area is where the subassemblies are manufactured. This area contains three cells which manufacture a single common subassembly and hence all three families can be manufactured in any of the three cells. The fabrication area has a conveyor system which transfers the products from one machine to another based on one-piece flow principle.

Operations in Fabrication Cells

There are three operations associated with the fabrication area:

- Lamination
- Slicing and Bottling
- Capping

The machines used for operation 1 in all three cells are similar and work under the same velocities (120 vials/min) but the number of machines within each cell varies. Operation 2 has machines that process 17 vials/min and 40 vials/min. Similarly, operation 3 has machines that process 78 vials/min and 123 vials/min. Table 2 shows the distribution of machines and velocities among the three cells.

Table 1. Product structures of families

	Components					
Family	S	L	I	B1	B2	B3
F1	1	1	1	1		
F2	2	1	1		1	
F3	4	1	1			1

Table 2. Number of machines and their production rates in fabrication cells

	Op. 1	Op.2		Op. 3		Output of Bottleneck (vials/min)
		Type I	Type II	Type I	Type II	
Production Rate (vials/min)	120	17	40	78	123	
Cell 1	1	2	2	0	1	114
Cell 2	1	4	0	1	0	68
Cell 3	2	3	2	0	2	131

Packaging Cells

The packaging area also has a conveyor system similar to the fabrication area which transfers products within packaging cells and also from the fabrication cells to the packaging cells. In the packaging area, the subassemblies produced in the fabrication area are used to produce the various finished products. The packaging cell 1 is semiautomatic while cells 2 and 3 are automatic. This difference in the types of machines results in constraints that do not allow the packaging of certain products in certain cells. There are a total of 36 finished products which differ in the quantity of vials they contain, the type of raw material the vials are made of, and the destination of the country to where they are shipped. The original cell feasibility matrix for the families is given in Table 3 and the restrictions are due to constraints in the packaging of the vials.

Table 3. Feasibility matrix of families and packaging cells

Family	Packaging Cell 1	Packaging Cell 2	Packaging Cell 3
F1	X	X	
F2	X	X	X
F3	X		X

Operations in Packaging Cells

There are five operations performed in packaging area and each operation requires one machine. The operations are described as follows:

- Feeding (This operation is only performed in the case of disconnected cells)
- Labeling
- Assembly (Automatic in cells 2 and 3, semi-automatic in cell 1)
- Sealing
- Bar Coding

Table 4 shows the production rates of the machines in all cells.

ALTERNATE DESIGNS CONSIDERED

In this section, the current product-cell feasibility restrictions are discussed for both connected and disconnected cellular systems.

Connected Cells

In this system, cells are set up such that the packaging cells form an extension or continuation of the respective fabrication cells. In other words, the output of a cell in fabrication area becomes the input for the corresponding packaging cell. Hence, it is referred to as a connected system. The connected system for the current product-

Table 4. Production rates for assembly-packaging machines in vials/minute

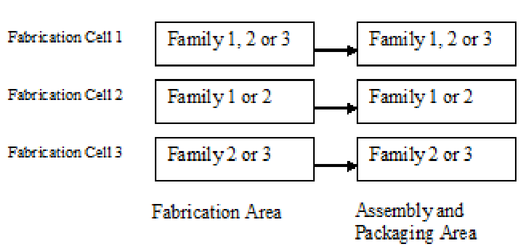
Cell	Family	Operation				
		4	5	6	7	8
Cell 1	Family 1	160	135	80	150	150
	Family 2	160	135	80	150	150
	Family 3	160	135	80	150	150
Cell 2	Family 1	160	135	100	150	150
	Family 2	160	135	180	150	150
	Family 3	NA	NA	NA	NA	NA
Cell 3	Family 1	NA	NA	NA	NA	NA
	Family 2	160	135	150	150	150
	Family 3	160	135	280	150	150

cell feasibility is shown in Figure 3. The output of family 1, family 2, and family 3 is essentially based on the bottleneck or the slowest machine in each cell of the fabrication or the packaging area and they are shown in Table 5.

Disconnected Cells

In this case, the output of a cell in the fabrication area can become an input for more than one cell in the packaging area depending upon the constraints in the packaging area. This can be considered to be a partially flexible disconnected cells type of system. The cell routing for each family is shown in Figure 4. In this figure, solid lines indicate that all the products processed in that particular fabrication cell can be processed in the assembly and packaging cell that they are connected to. On the other hand, the dashed lines show that only some of the products processed in the fabrication cell can be processed in the corresponding assembly and packaging cell. This provides a greater amount of flexibility with respect to the routing of the parts in the cellular system. The output rates of family 1, family 2, and family 3 depend on the fabrication-packaging cell combination and they are determined by the slowest machine as shown in Table 6.

Figure 3. Cell routing of families for the connected system



Cases Considered

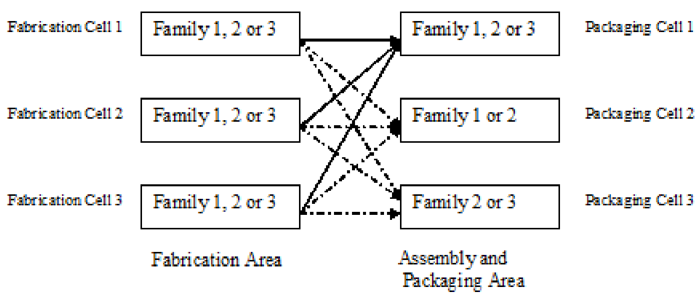
The experimentation discussed in this chapter can be grouped in the following sections:

- Original Family-Cell Feasibility Matrix
  Production orders are based on customer orders.
- Various Family-Cell Feasibility Options
  Seven different family-cell feasibility options have been considered as given in Table 7. In this case too, production orders are based on customer orders.

Table 5. Output rates for cells in the connected system

Cell #	Family #	Output Rate of the Bottleneck Machine in Fabrication Area (vials/min)	Output Rate of the Bottleneck Machine/Operator in Packaging Area (vials/min)	Output Rate (vials/min)
Cell 1	Family 1	114	80	80
	Family 2		80	80
	Family 3		80	80
Cell 2	Family 1	68	100	68
	Family 2		135	
Cell 3	Family 2	131	135	131
	Family 3		135	

Figure 4. Cell routing of families for disconnected system



METHODOLOGY USED

This section describes the methodology used to develop the different simulation models in Arena 7.0.

Input Data Analysis

Input data such as customer order distributions, their respective inter-arrival times, processing times, and routings were all obtained based on the data provided by the company. The data provided was basically the total sales volume in vials for each part belonging to one of the three families for a period of nineteen months. Table 8 shows the customer order sizes and the inter-arrival time distributions for each product.

Simulation Models

The models were run 24 hours a day which basically represented 3 shifts round the clock. Setup times and material handling times were negligible. Preemption was not allowed due to material control restrictions by FDA. Vials move based on one-piece flow between machines. The simulation models are discussed for different cases separately in the following paragraphs.

Case 1: Connected Cells: After the entities are created, they are routed to cells 1, 2 or 3 based on the type of family they belong to. The entities enter the fabrication area as a batch equivalent to the customer order size. Once a batch of entities enters the cell they are split and there is a one-piece flow in the cell. Entities belonging to a family go to one of its feasible cells based on the shorter queue length among 2nd operation. This is done

Table 6. Output rate of each routing combination for the disconnected system

Family #	Fabrication Area Cell (Output of the Bottleneck Machine in vials/min)	Packaging Area Cell (Output of the Bottleneck Machine in vials/min)	Output Rate of Routing Combination (vials/min)
Family 1	Cell 1 (114)	Cell 1 (80)	80
	Cell 1 (114)	Cell 2 (100)	100
	Cell 2 (68)	Cell 1 (80)	68
	Cell 2 (68)	Cell 2 (100)	68
	Cell 3 (131)	Cell 1 (80)	80
	Cell 3 (131)	Cell 2 (100)	100
Family 2	Cell 1 (114)	Cell 1 (80)	80
	Cell 1 (114)	Cell 2 (135)	114
	Cell 1 (114)	Cell 3 (135)	114
	Cell 2 (68)	Cell 1 (80)	68
	Cell 2 (68)	Cell 2 (135)	68
	Cell 2 (68)	Cell 3 (135)	68
	Cell 3 (131)	Cell 1 (80)	80
	Cell 3 (131)	Cell 2 (135)	131
Family 3	Cell 1 (114)	Cell 1 (80)	80
	Cell 1 (114)	Cell 3 (135)	114
	Cell 2 (68)	Cell 1 (80)	68
	Cell 2 (68)	Cell 3 (135)	68
	Cell 3 (131)	Cell 1 (80)	80
	Cell 3 (131)	Cell 3 (135)	131

Table 7. Different family-cell feasibility options

Cellular System	Cell Type	Cell	Options						
			O1	O2	O3	O4	O5	O6	O7
Connected Cells	Fab. Cells	C1	1	1,2,3	1,2	1,2,3	1	1,2	1,2
		C2	2	1,2,3	2,3	1,2	2	2,3	2,3
		C3	3	1,2,3	1,3	2,3	3	1,3	1,3
	Pack. Cells	C1	1	1,2,3	1,2	1,2,3	1	1,2	1,2
		C2	2	1,2,3	2,3	1,2	2	2,3	2,3
		C3	3	1,2,3	1,3	2,3	3	1,3	1,3
Disconnected Cells	Fab. Cells	C1	1	1,2,3	1,2	1,2,3	1	1,2,3	1,2
		C2	2	1,2,3	2,3	1,2	2	1,2,3	2,3
		C3	3	1,2,3	1,3	2,3	3	1,2,3	1,3
	Pack. Cells	C1	1,2,3	1,2,3	1,2,3	1,2,3	1	1,2	1,2
		C2	1,2,3	1,2,3	1,2,3	1,2,3	2	2,3	2,3
		C3	1,2,3	1,2,3	1,2,3	1,2,3	3	1,3	1,3

Table 8. Inter-arrival time and customer order size distributions for products

Family #	Product #	Inter-arrival Time Distribution	Customer Order Size Distribution
Family 1	1	0.999 + WEIB(0.115, 0.54)	1.09 + LOGN(1.56, 1.06)
	2	0.999 + WEIB(0.0448, 0.512)	TRIA(18, 23.7, 52)
	3	1.11 + EXPO(1.87)	9 + WEIB(7.66, 1.27)
	4	2 + LOGN(3.19, 3.68)	2 + 17 * BETA(0.387, 0.651)
	5	4 + LOGN(5.05, 14)	207 + LOGN(86.5, 139)
	6	UNIF(0, 26)	TRIA(6, 12.5, 71)
	7	-0.001 + 26 * BETA(0.564, 0.304)	UNIF(9, 80)
	8	TRIA(0, 6.9, 23)	EXPO(25.3)
	9	NORM(13.7, 7.49)	NORM(108, 30.8)
	10	6 + WEIB(3.78, 0.738)	TRIA(98, 120, 187)
	11	UNIF(0, 26)	UNIF(14, 34)
Family 2	12	0.999 + WEIB(0.0126, 0.405)	5 + WEIB(7.51, 0.678)
	13	1 + LOGN(0.99, 2.62)	2 + 11 * BETA(0.412, 0.527)
	14	1.24 + EXPO(1.46)	30 + 26 * BETA(0.643, 1.08)
	15	EXPO(7.06)	2 + 34 * BETA(0.321, 0.519)
	16	0.999 + WEIB(0.0313, 0.503)	NORM(149, 57.1)
	17	0.999 + WEIB(0.195, 1.12)	NORM(23, 14.2)
	18	TRIA(0, 11.2, 25)	101 * BETA(0.822, 0.714)
	19	26 * BETA(0.649, 0.42)	EXPO(154)
	20	EXPO(7.4)	UNIF(0, 90)
	21	UNIF(0, 26)	TRIA(0, 231, 330)
	22	28 * BETA(1.11, 0.547)	TRIA(0, 224, 325)
	23	27 * BETA(0.679, 0.429)	EXPO(119)
	24	28 * BETA(0.468, 0.255)	TRIA(425, 1.05e+003, 2.5e+003)
	25	1.16 + LOGN(2.48, 1.76)	NORM(867, 534)
	26	EXPO(7.03)	NORM(68, 32.8)
	27	TRIA(0, 4.44, 25)	EXPO(13.8)
	28	9 + 17 * BETA(0.559, 0.0833)	24 * BETA(0.67, 0.969)
	29	28 * BETA(0.466, 0.301)	NORM(420, 168)
	30	28 * BETA(0.932, 0.479)	NORM(267, 110)
	31	2 + 26 * BETA(0.314, 0.458)	TRIA(0, 274, 381)
	32	UNIF(0, 26)	TRIA(0, 297, 368)
Family 3	33	0.999 + WEIB(0.0117, 0.424)	TRIA(843, 1.19e+003, 2e+003)
	34	1.33 + 1.96 * BETA(0.3, 0.636)	WEIB(6.83, 0.613)
	35	1 + LOGN(5.23, 7.03)	37 + LOGN(147, 1.51e+003)
	36	4 + 22 * BETA(0.305, 0.197)	TRIA(0, 543, 591)

because the second operation in each cell has been identified as the bottleneck operation based on trial runs conducted. In cell 1 and cell 3, the entities undergo operation 1 and go to operation 2 where there are two types of machines namely the slow (Type I) and fast (Type II) machines available for processing. The entities are routed to either type of machine based on a percentage which was decided after a number of simulation runs in order to minimize the queue lengths and hence the waiting time. In cell 1, 30% of the entities were routed to the Type I machine and the rest were routed to the Type II machine. In cell 3, 40% of the entities were routed to the Type I machine and the rest were routed to the Type II machine.

Each of the entities leaving the fabrication cells enters the corresponding packaging cells. For example, entities from cell 1 in the fabrication area will enter cell 1 of the packaging area. The entities entering the packaging area undergo processing through operation 4. In the fifth operation, the vials are grouped based on the type of family they belong to. Family 1 consists of only 1 vial, family 2 consists of 2 vials and family 3 consists of 4 vials. Thus, the vials that are batched in Arena after operation 5 are processed in operations 6, 7 and 8 where they are boxed, sealed and coded. In the final batching, the vials are batched together in a box based on the final customer order sizes. The final batch sizes are the same as the input batch sizes. There is a waiting time associated since the entities might have to wait till the required batch size is reached and only then get disposed.

The warm up time for the model was determined to be 2000 hours based on steady state analysis. The simulation was run for 2500 hours after the end of the warm-up period.

Case 1: Disconnected Cells: The entities enter the fabrication area in batches as explained for the connected system. The batches of entities in disconnected system are routed differently as compared to the connected system. Here, the batches of entities are routed to cell 1, cell 2, or cell 3 of the fabrication area based on the shortest queue length of the bottleneck operation which is operation 2 as explained earlier. The flexibility of routing the families to any of the cells in this type of system is the only major difference between the connected and disconnected systems in the fabrication area. The processing times of the machines and the sequence of operations for the entities for both systems are the same. Since the flow is disconnected in this system, the entities are batched again to the same customer order sizes at the end of the fabrication area.

The batches of entities entering the packaging area are routed to specific packaging cells based on shortest queue length as shown earlier in Table 4. These batches are then split and the entities follow a one-piece flow. Also, there is an extra feeding operation at the start of the packaging cells in order to accommodate the transfer of entities from fabrication to packaging. The method in which the entities are transferred from fabrication to packaging and the extra feeding operation is the only major difference between the connected and disconnected systems in the packaging area. The processing times of the machines and the sequence of operations for the entities for both systems are the same.

Case 2: It is very similar to case 1 except that the routings for products are varied as given in Table 7. In this table, Option 5 (O5) is the least flexible arrangement where each cell can process only one product family for both connected and disconnected cells. Option 2 (O2) is the most flexible arrangement with three cells capable of running all three product families both in connected cells and disconnected cells. The remaining options vary in flexibility between O5 and O2. In Option 1, the system is highly inflexible in connected cells whereas it is very flexible in packaging cells of disconnected arrangement (three product families for each cell). In options 3, 4, 6 and 7, each product family can be run at least in two cells. In option 3, packaging cells of disconnected arrangement is more flexible (once again three product families for each cell). In option 4, a little bit more flexibility is added to both connected and disconnected cells (cell 1 can run three families). In option 6, more flexibility is now added to fabrication cells of disconnected system (three product families for each cell). In option 7, each family can be run in two cells. However, models for options 1 and 5 didn't stabilize and therefore they were not included in comparisons.

Production order quantities for products 33 and 36 were both reduced by 40% and 50%, respectively to fit into existing capacity for case 1. Validation and verification are an inherent part of any computer simulation analysis. Models were verified and validated before statistical analysis was performed for all scenarios.

RESULTS OBTAINED

The results obtained from simulation analysis for average flowtime and average work-in-process inventory are summarized in Tables 9 and 10, respectively. The results are based on 100 replications. The statistical analysis was conducted using the statistical functions available in Excel. A t-test assuming unequal variances for two samples was conducted for a 95% confidence interval for each family under each system. Table 11 displays the comparison for each family with respect to flowtimes and work-in-process between connected and disconnected systems. Table 12 displays comparisons for the families for the same performance measures but the comparisons are made between different connected systems from cases 1 and 2. Table 13 also displays comparisons for the families for the same performance measures but the comparisons are made between different disconnected systems from cases 1 and 2. Results are denoted as significant (S) or not significant (NS) based on the conclusions reached. Also whenever significant, better option was denoted in a parenthesis. The significance of the results was based on the p-value obtained from the T-test conducted for an alpha level of 0.05. As mentioned earlier, no results for options 1 and 5 were obtained as the system did not stabilize.

As observed in Table 11, for case 1, the flowtimes and work-in-process were observed to be different and the disconnected system had lower flowtimes and WIP for families F1 and F3 while the difference was significant for F1. On the other hand, WIP was significantly lower for F2 in the connected system. For case 2 with all the options considered, when there was a significant difference, this was always in favor of connected systems. For option 2, the flowtime for family 2 and the WIP for all three families for the connected system were significantly lower than those of in the disconnected system. For options 3, 6, and 7 which were the same for the connected system, the flowtimes and WIP for families 1 and 2 were significantly lower than the disconnected

Table 9. Average flowtime results for all cases

Cases and Options	Connected Cells Configuration			Disconnected Cells Configuration		
	F1	F2	F3	F1	F2	F3
C1	42.66	50.52	87.53	31.19	54.39	71.61
C2-02	31.08	45.98	66.61	32.55	51.62	73.79
C2-03	24.91	39.84	67.06	27.24	46.93	83.48
C2-04	41.26	51.15	78.25	35.14	49.66	79.49
C2-06	Same as C2-03			31.88	51.17	73.80
C2-07	Same as C2-03			70.67	45.91	78.06

Table 10. Average work-in-process results for all cases

Cases and Options	Connected Cells Configuration			Disconnected Cells Configuration		
	F1	F2	F3	F1	F2	F3
C1	128.59	1403.77	1381.40	100.15	1622.52	1182.29
C2-02	90.00	1184.19	1052.06	99.70	1563.94	1267.13
C2-03	70.67	1046.90	1246.10	86.36	1425.42	1442.67
C2-04	126.10	1425.71	1269.42	111.27	1667.29	1409.31
C2-06	Same as C2-03			97.46	1555.34	1273.61
C2-07	Same as C2-03			80.34	1380.77	1532.79

Table 11. Connected vs. disconnected configuration for each family

Cases and Options	FLOWTIME			WIP		
	F1	F2	F3	F1	F2	F3
C1	S (D)	NS	NS	S (D)	S (C)	NS
C2 – O2	NS	S (C)	NS	S (C)	S (C)	S (C)
C2 – O3	S (C)	S (C)	S (C)	S (C)	S (C)	NS
C2 – O4	NS	NS	NS	NS	S (C)	NS
C2 – O6	S (C)	S (C)	NS	S (C)	S (C)	NS
C2 – O7	S (C)	S (C)	NS	S (C)	S (C)	NS

Table 12. Comparison between connected systems

Cases and Options	FLOWTIME			WIP		
	F1	F2	F3	F1	F2	F3
O2 VS O3	S (O2)	S (O2)	NS	S (O2)	S (O2)	NS
O2 VS O4	S (O2)	NS	NS	S (O2)	S (O2)	NS
O3 VS O4	S (O3)	S (O3)	NS	S (O3)	S (O3)	NS
C1 VS O2	S (O2)	NS	NS	S (O2)	S (O2)	NS
C1 VS O3, O6,O7	S (O3)	S (O3)	NS	S (O3)	S (O3)	NS
C1 VS O4	NS	NS	NS	NS	NS	NS

system. For option 4, the WIP for family 2 in the connected system was the only significant result. From Table 12, it can be observed that option 2 (O2) provided the best results when compared to rest of the options within the connected system with lower flowtimes and WIP followed by option 3 (O3). From Table 13, it can be observed that the flowtimes and WIP for options 3 and 7 (O3, O7) were consistently and significantly better when compared to the rest of the options in the disconnected cells configuration. Also, when these two options were compared against each other there was no significant difference observed for any of the families and performance measures. A comparison between models C1 and O2 did not yield any significant results either and were definitely

Table 13. Summary table of results for disconnected system: cases 1 and 2

Cases and Options	FLOWTIME			WIP		
	F1	F2	F3	F1	F2	F3
O2 VS O3	S (O3)	S (O3)	NS	S (O3)	S (O3)	NS
O2 VS O4	S (O4)	S (O4)	NS	S (O4)	S (O4)	NS
O2 VS O6	S (O6)	S (O6)	NS	S (O6)	S (O6)	NS
O2 VS O7	S (O7)	S (O7)	NS	S (O7)	S (O7)	NS
O3 VS O4	S (O3)	NS	NS	S (O3)	S (O3)	NS
O3 VS O6	S (O3)	S (O3)	NS	S (O3)	S (O3)	NS
O3 VS O7	NS	NS	NS	NS	NS	NS
O4 VS O6	NS	NS	NS	S (O6)	NS	NS
O4 VS O7	S (O7)	S (O7)	NS	S (O7)	S (O7)	NS
C1 VS O2	NS	NS	NS	NS	NS	NS
C1 VS O3	S (O3)	S (O3)	NS	S (O3)	S (O3)	S (C1)
C1 VS O4	NS	NS	NS	NS	NS	NS
C1 VS O6	NS	NS	NS	NS	NS	NS
C1 VS O7	S (O7)	S (O7)	NS	S (O7)	S (O7)	NS

less superior in performance when compared with the rest of the options.

CONCLUSION

In this chapter, the performance of connected and disconnected cellular systems was compared under make-to-order strategy in a real cellular setting. In the existing system (case 1), it was observed that no cellular manufacturing design dominated the other, i.e., mixed results were obtained as to which system did better for each family. The flowtime and work-in-process for family 1 for the disconnected system were lower. On the other hand, the WIP for family 2 in the connected system was lower. The other comparisons did not yield any significant results and hence dominance could not be established in terms of better cellular system.

In case 2, which is basically an extension of case 1, the impact of considering alternate cell routings for each part family was studied for both connected cells and disconnected cells. In most cases, connected cells outperformed disconnected cells with respect to both average flowtime and WIP, especially for family 1 and family 2. This leads to the conclusion that the connected system is the better system in this situation since family 1 and family 2 make up for 32 of the 36 products and comprise of about 85% by volume of the production orders in the system. The average flowtime and WIP conclusions are similar but not identical, i.e. there were incidents where flowtime was significantly better but not necessarily corresponding WIP and vice versa. If one wanted to choose the best connected cell configuration, that would be option 2. This is possibly due to option 2 having the highest flexibility among all options as each family could be routed to any of the fabrication and packaging cells. Options 3, 4 and case 1 followed in the order of performance leading to the conclusion that increase in routing flexibility of the families resulted in significantly lower flowtimes and WIP.

A similar comparison among all options developed for the disconnected system showed that options 3 and 7 performed better than the rest of the options. Option 3 had complete flexibility in the packaging area but limited flexibility in the fabrication area and option 7 had limited flexibilities in both the areas. Limited flexibility as applicable to these two options means that each family could go to at least two specified cells. On the other hand, option 2 was the worst performing system among the options for case 2 even though it had the highest flexibility. This can be attributed to the fact that routing decisions are made based on queue sizes only. Family 3 products have the highest processing times and it is possible that queues in all cells may contain products from family 3 thus leading to higher lead times for the parts that join that queue.

For case 1 and also option 2 from case 2, the disconnected system was modified to delete the extra feeding operation and the batching at the end of the fabrication area. This was done in order to determine the reason why the connected system performed better than the disconnected system in most of the comparisons made. The two modified simulation models were run and the results were statistically analyzed. In case 1, the flowtime for family 1 and the WIP for family 2 was significantly better for the disconnected system. In the original comparison, WIP and flowtime for family 1 in the disconnected system was better and the WIP for family 2 in the connected system was significantly better. The rest of the comparisons did not yield any significant results. For option 2, none of the comparisons yielded significant results as opposed to the original comparison when the connected system clearly performed better than the disconnected system. From these results it can be concluded that the extra operation and the extra batching increases the average WIP and flowtimes for each of the families and could be responsible for the disconnected system not performing as well as or better than the connected system.

REFERENCES

Agarwal, A., & Sarkis, J. (1998). A review and analysis of comparative performance studies on functional and cellular layouts. *Computers & Industrial Engineering*, *34*(1), 77–89. doi:10.1016/S0360-8352(97)00152-6

Albino, V., & Garavelli, C. A. (1998). Some effects of flexibility and dependability on cellular manufacturing system performance. *Computers & Industrial Engineering*, *35*(3-4), 491–494. doi:10.1016/S0360-8352(98)00141-7

Caprihan, R., & Wadhwa, S. (1997). Impact of routing flexibility on the performance of an FMS – A simulation study. *International Journal of Flexible Manufacturing Systems*, *9*, 273–278. doi:10.1023/A:1007917429815

DeCroix, G. A., & Arreola-Risa, A. (1998). Make-to-order versus make-to-stock in a production inventory system with general production times. *IIE Transactions*, *30*, 705–713. doi:10.1023/A:1007591722985

Federgruen, A., & Katalan, Z. (1999). The impact of adding a make-to-order item to a make-to-order production system. *Management Science*, *45*(7), 980–994. doi:10.1287/mnsc.45.7.980

Flynn, B. B., & Jacobs, F. R. (1987). A comparison of group technology and process layout using a model of an actual shop. *Decision Sciences*, *18*, 289–293.

Gupta, D., & Benjaafar, S. (2004). Make-to-order, make-to-stock, or delay product differentiation? A common framework for modeling and analysis. *IIE Transactions*, *36*, 529–546. doi:10.1080/07408170490438519

Johnson, J., & Wemmerlov, U. (1996). On the relative performance of functional and cellular layouts – An analysis of the model-based comparative studies literature. *Production and Operations Management*, *5*(4), 309–334. doi:10.1111/j.1937-5956.1996.tb00403.x

Morris, S. J., & Tersine, J. R. (1990). A simulation analysis of factors influencing the attractiveness of group technology cellular layouts. *Management Science*, *36*(12), 1567–1578. doi:10.1287/mnsc.36.12.1567

Nandi, A., & Rogers, P. (2003). Behavior of an order release mechanism in a make-to-order manufacturing system with selected order acceptance. *Proceedings of the 2003 Winter Simulation Conference.*

Selen, J. W., & Ashayeri, J. (2001). Manufacturing cell performance improvement: A simulation study. *Robotics and Computer-integrated Manufacturing*, *17*, 169–176. doi:10.1016/S0736-5845(00)00051-X

Shafer, S. M., & Charnes, J. M. (1993). Cellular versus functional layouts under a variety of shop operating conditions. *Decision Sciences*, *24*(3), 665–681. doi:10.1111/j.1540-5915.1993.tb01297.x

Süer, G. A., Huang, J., & Maddisetty, S. (2010). Design of dedicated, shared and remainder cells in a probabilistic environment. *International Journal of Production Research*, *48*(19), 5613–5646. doi:10.1080/00207540903117865

Van Donk, D. P. (2000). Make to stock or make to order: The decoupling point in the food processing industries. *International Journal of Production Economics*, *69*, 297–306. doi:10.1016/S0925-5273(00)00035-9

Yazici, J. H. (2005). Influence of flexibilities on manufacturing cells for faster delivery using simulation. *Journal of Manufacturing Technology*, *16*(8), 825–841. doi:10.1108/17410380510627843

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 37-52, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 39
AutomatL@bs Consortium:
A Spanish Network of Web-based Labs for Control Engineering Education

Sebastián Dormido
Universidad Nacional de Educación a Distancia, Spain

Héctor Vargas
Pontificia Universidad Católica de Valparaíso, Chile

José Sánchez
Universidad Nacional de Educación a Distancia, Spain

ABSTRACT

This chapter describes the effort of a group of Spanish universities to unify recent work on the use of Web-based technologies in teaching and learning engineering topics. The network was intended to be a space where students and educators could interact and collaborate with each other as well as a meeting space for different research groups working on these subjects. The solution adopted in this chapter goes one step beyond the typical scenario of Web-based labs in engineering education (where research-groups demonstrate their engineering designs in an isolated fashion) by sharing the experimentation resources provided by different research groups that participated in this network. Finally, this work highlights the key points of this project and provides some remarks about the future use of Web-based technologies in school environments.

INTRODUCTION

The evolution of the Internet has changed the education landscape drastically (Bourne et al. 2005, Rosen 2007). What was once considered distance education is now called online education. In other words, the method of teaching and learning is based on the use of the Internet to complete educational activities. A specific example of this new teaching model is the Spanish University for Distance Education (UNED). Compared to other Spanish universities, this institution has the largest number of students because distance education allows students to obtain a degree or improve

DOI: 10.4018/978-1-4666-1945-6.ch039

their professional skills without having to change their lifestyles. UNED is not a unique institution; there are many universities around the world with an online presence, such as the Open University in Colombia, Open Universities in Australia, the Open University in UK, the Open University in Catalonian, Fern Universität in Germany, and many more. The existence of these institutions confirms the viability and importance of computer-assisted teaching and learning through the Internet.

The implementation of a distance learning model is not an easy task in engineering and sciences studies (Williams 2007). In addition to textual/multimedia information and other resources required to demonstrate theoretical aspects in an online course, hands-on laboratories should also be included. This requirement is particularly necessary for control engineering, which is an inherently interdisciplinary field in which progress is achieved through a mix of mathematics, modeling, computation, and experimentation (Astrom 2006). In this context, students should be able to

- Understand the underlying scientific model of the phenomenon that was studied.
- Become acquainted with the limits of the model (i.e., how does the model accurately reflects real behavior and to what extent it remains a basic approximation).
- Learn how to manipulate the parameters of the model in order to fine-tune the behavior of the real system. (Dormido 2004)

To achieve these goals, the implementation of an effective Web-based educational environment for any engineering topic should cover three aspects of the technical education: concept, interpretation, and operation. The student should be provided with an opportunity to become an active player in the learning process (Dormido et al. 2005). In this context, the potential for Web-based experimental applications such as virtual laboratories (Valera et al. 2005), remote laboratories (Casini et al. 2004, Brito et al. 2009) and games (Eikaas et al. 2006) as pedagogical support tools in the learning/teaching of control engineering has been presented in many works. In fact, in the last decade, several academic institutions have explored the World Wide Web (WWW) to develop their courses and experimental activities in a distributed context. However, most of these developments have focused only on the technical issues related to building Web-enabled applications for performing practical activities through the Internet (e.g., how to start up remote monitoring of a real device or how to build sophisticated virtual interfaces). At most, these implementations may include a set of Web pages with a list of activities that need to be carried out by the users. Some examples of these implementations are provided in the additional reading section at the end of the chapter. In general, these developments do not take into account the social context of the interactions and the collaboration that is typically generated in traditional hands-on laboratories (Nguyen 2007). Indeed, direct contact with teachers and interactions with classmates are valuable resources that may be reduced or even disappear when hands-on experimental sessions are conducted via Web-based laboratories.

New trends in the use of Web-based resources for teaching and learning in the engineering disciplines include the use of Web 2.0 technologies such as social software in building virtual representations of face-to-face (f2f for short) laboratories in a networked, distributed environment (Gillet et al., 2009). This objective was grounded in the idea that educational institutions and many workplaces are equipped with a type of tool that connects people, contents and learning activities and can thus transfer information and knowledge. *Learning to learn* is the new challenge for the new generation of students. In other words, they have to learn to use Web resources to improve their teaching and learning.

Commonly, a mix of Web-based technologies and software agents (Salzmann & Gillet 2008) is used to develop remote experimentation systems for pedagogical purposes. For this reason, most of the remote experimentation systems are custom-made solutions. This means that the selection of software tools and global system architecture are not simple tasks due to the wide variety of software frameworks that are available. This chapter describes the structure of the remote experimentation system used in this study, which is based on the use of three software tools: Easy Java Simulations (Easy Java 2010), LabVIEW (LabVIEW 2010), and eMersion (eMersion 2010).

BACKGROUND

In a typical scenario for remote experimentation, universities provide the overall infrastructure required for the remote experimentation services offered to students, including a set of didactic setups that are specially designed for hands-on laboratories, a set of server computers used to interface these processes and a main server computer providing the complementary Web-based resources necessary to use the remote labs. The system users (clients) can access experimentation services from any Internet connection. However, developing a complete environment for experimentation services is not an easy task. For this reason, this section presents a systematic approach for developing such systems.

Although the development of an application with the previously described features can be structured in multiple ways, we have divided the problem into two levels or "layers." The first layer is the *experimentation layer*, which includes all the necessary software and hardware components needed to develop the experimental applications for the Web-based virtual or remote laboratories. Because Web-based labs do not supply all the elements needed to provide remote experimentation services, complementary Web-based resources are needed to manage students' learning. Thus, the *e-learning layer* incorporates the development of functionalities required to support teaching and learning through the Internet (Vargas et al., 2008).

Layer 1: The Experimentation Layer

The experimentation layer included the design methodology and construction of a graphical user interface for clients as well as the server application. This layer was developed with a client and server structure. The first step to start the development process was the analysis of the requirements and specifications for implementing the interface. The following subsections include some recommendations for developing this layer.

Requirements and Specifications for the Client

- The software should be multiplatform. For example, Java has the required characteristics for designing this kind of application; the user only needs a Web browser with Java support to access the Web-based lab.
- The protocol to communicate with the server should include low-level protocols to stream data through the Internet, such as TCP (Transfer Control Protocol) or UDP (User Datagram Protocol). These protocols allow for better control of data packet transmissions in networks.
- The graphical user interface must be simple and intuitive. It should be user-friendly and useful in different environments.
- Either virtual or remote access to the laboratory should be enabled with the same graphical interface. In simulation mode, the state of the system and its associated variables must be updated based on the evolution of a mathematical model of the process. Otherwise, in remote mode, these variables should be updated according to the real plant changes in the remote location. Video feedback should also be includ-

ed to provide distance users with a sense of presence in the laboratory.

- Events scheduling for program faults in the system should be included. The systems could be enabled to analyze systems in the presence of noise or disturbance measurements. The robustness of the system could also be evaluated under anomalous operating situations.
- Finally, it is recommended that the users define experiments in an easy manner. For example, a programmed change of a set-point value could be required to observe the process response at different operation points.

Requirements and Specifications for the Server

From a software design point of view, the server is composed of a set of modules (some of them optional) that are described below:

- A data exchange module: this software remains in a listening state while it is waiting for remote connections from users. It receives commands and queries from clients and makes these inputs effective over a physical system. The responses are retrieved from the real plant through the instrumentation hardware and are sent back to the client. The link between the server and clients would be established via a TCP/IP protocol suite.
- An access management module: this software module would manage all of the information related to the management of users and timeslot bookings for the use of real plants. A database manager would be used to handle the users' reservations and physical resources. Each register in this database would correspond to a booking scheduled for a specific time and date.
- An instrumentation module: this module would incorporate all of the hardware needed to connect the physical system with the server.
- A remote visualization module: this module would allow users to examine what is happening with the physical system during its remote manipulation by a client. Video cameras would be used to facilitate this feature. This system must be able to transmit a sense of realism to encourage its usage and increase the motivation of the users while completing tasks.
- Each module in the server has a counterpart on the client side. For example, to read the video streaming captured by the server, the client application must implement a software module that allows for the retrieval and decoding of video streaming from the remote camera and then renders it in the interface.

Layer 2: The E-Learning Layer

The previous section presented the main features that need to be taken into account when analyzing, designing, and building the experimentation layer. Additionally, a second key aspect that should be addressed is the development and/or use of a Web-based learning management system (LMS) to support a student's learning process. This platform should organize user access to the experimentation modules that are available and allow for students/teachers to interact and collaborate with one another. The implementation phase would require the following:

- Simplification of the organization of user groups.
- Notification services by email, instant messaging, news, and other methods.
- Documentation such as practical guides, task protocols, instruction manuals and any other information needed to per-

form a remote experimentation session autonomously.

- A sequence of activities that students must carry out during an experimental session. There can be two types of tasks: 1) tasks in *simulation mode* and 2) tasks in *remote mode*. Tasks in simulation mode are tasks that students must carry out prior to performing the experiments in the real plant. These tasks should be completed with a graphical user interface that allows students to work in a simulated environment. The objective should be to gain adequate insight about the procedures that are involved in the experiment. These tasks would reduce the time students spend on activities using the real plant. Remote access should not be allowed if the student has not satisfactorily completed the required tasks in simulation mode. If the student's work was evaluated positively by the teaching staff, then access to remote mode would be granted.
- A method for managing students and the assessment of their work as well as uploading reports.
- An automatic booking system to schedule access to the physical resources.
- At the end of the development process, the *experimentation* and *e-learning* layers have to be integrated to produce the final Web application for virtual and remote labs. This integration needs to establish certain links and channels between the Web modules from both layers. For example, in our particular framework, we made it possible to save data collected from an experimentation applet (experimentation layer) in a shared Web space that was part of the e-learning layer. The data stored in this space could be retrieved later for analysis.

IMPLEMENTATION

The implementation process of the remote experimentation system that is described in this chapter can be itemized into two independent processes that were combined to create the final environment. These developments can be summarized as follows:

- *Building hybrid laboratories for pedagogical purposes (the experimentation layer).*
- *Integrating the hybrid laboratories into a Learning Management System (LMS) to publish resources and provide mechanisms for accessing the real plants (the e-learning layer).*

Building Hybrid Laboratories for Pedagogical Purposes

A *hybrid laboratory* provides remote software simulations and real experiments in a single environment that can be accessed over the Internet. The client/server approach is commonly used for the technical implementation of both features (Callaghan et al. 2006, Zutin et al. 2008). Specifically, when a student is conducting an experiment in a virtual manner, he or she also works with a mathematical model of a process.

When developing the simulated portion of a *hybrid laboratory*, developers not only need to create a technology that covers all the aspects related to the use of simulations in a local mode. The applications must also work well in a distributed environment. The graphical user interface, for instance, could be a pure HTML/JavaScript application, or it could require a plug-in such as Flash, Java or ActiveX that runs in a Web browser.

Although one of the most relevant features of Java is the simplicity of the language, creating a graphical simulation in this programming language is not a straightforward task. Conceiving relatively complex Web-based applications requires advanced knowledge of object-oriented programming and other features of Java (Esquembre 2005). For this reason, the following subsection presents Easy Java Simulations (EJS), a software tool that was used to create the client interfaces for the hybrid laboratories in this chapter.

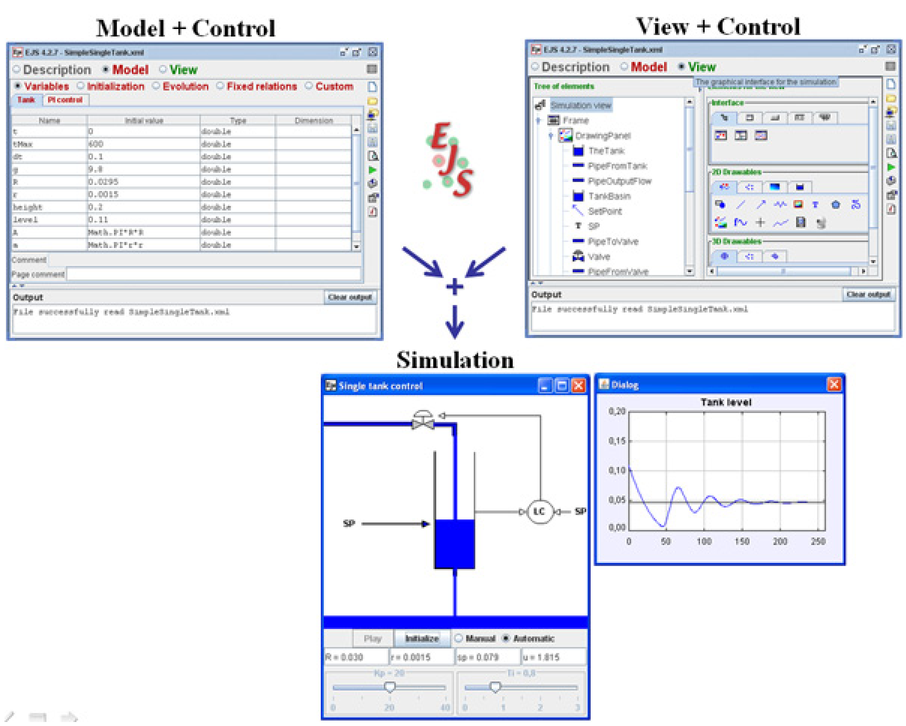
EJS as a Development Tool for Hybrid Laboratories

EJS is a freeware, open-source tool that was developed in Java and is specially designed for the creation of discrete computer simulations (Christian & Esquembre 2007). EJS was originally designed for users with little programming experience. However, users need to know the analytical model of the process and the design of the graphical interface in detail. The architecture of EJS was derived from the *model-view-control* (MVC) paradigm, a philosophy that is based on the fact that interactive simulations must include three parts:

- The *model*, which describes the process under study in terms of (1) variables, which define the different possible states of the process, and (2) the relationships between these variables, which are expressed by computer algorithms.
- The *control*, which defines certain actions a user can perform on the simulation.
- The *view*, which shows a graphical representation (either realistic or schematic) of the process states.

EJS makes programming simple by eliminating the control element of the MVC paradigm and fusing one part in the *view* and the other part in the *model*, as shown in Figure 1. Thus, applications can be created in two steps: (1) defining the *model* to simulate with the built-in simulation mechanism of EJS and (2) building the *view* by showing the model state and incorporating the changes made by users. Figure 1 shows a simple virtual-lab created

Figure 1. MVC paradigm abstraction of EJS



by EJS for teaching basic control concepts based on the well-known *single-tank process*.

Although EJS was initially conceived as a software tool to create interactive simulations for teaching physics, it has been successfully applied to many other research areas, including physical systems, mechanical systems, control systems (as in this essay), and medical systems. Thus, EJS could be classified as a general purpose tool intended to create interactive simulations of scientific phenomena based on models.

Finally, one of the most important features of EJS is that its applications can be easily distributed through the Internet in applet form. Applets are Java programs that can be executed in the context of a Web browser in a way similar to Flash or HTML/JavaScript applications. To find more information about the EJS mechanism, creating simulations, and additional features, please visit the EJS homepage (http://www.um.es/fem/EjsWiki/).

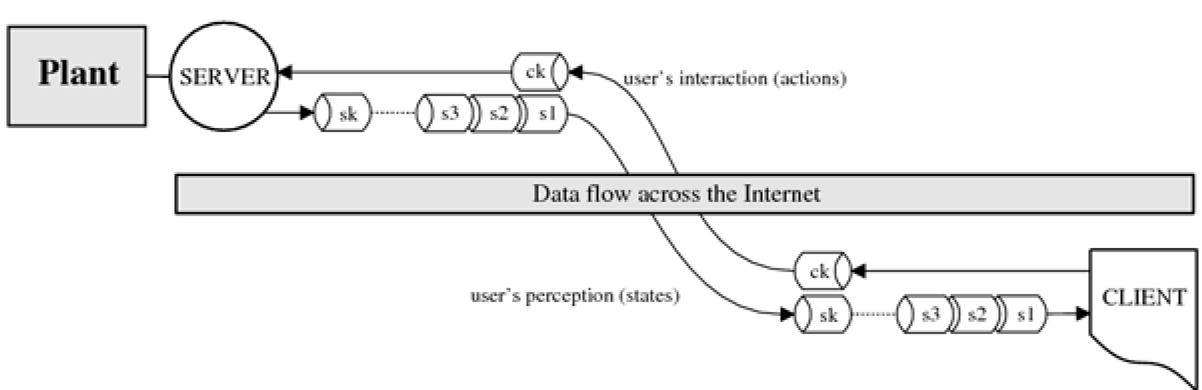
LabVIEW Server for Developing Hybrid Laboratories

Carrying out the remote operation of any physical device is challenging if we take into account the number of technical aspects that must be solved. In most cases, technical issues such as *performance*, *interaction level*, *visual feedback*, *real-time control*, *user perception*, *safety* and *fault tolerance* require extensive research (Salzmann & Gillet 2002).

In general, remote experimentation through the Internet requires an awareness of the current state of the distant real plant to change the value of any input parameter of the remote system and to perceive the effect of this change with a minimal transmission delay. Figure 2 shows the process in which a client application maintains a connection with a remote server to control a real plant remotely. The server-side sends a continuous flow of information, which is represented by the information blocks *s(k)* that reflect the current state of the plant, and the server receives information blocks *c(k)* containing changes in system input parameters that were carried out by a remote user. The client side of the system receives information on the state of the system that is sent by the server (contained in *s(k)* blocks) while simultaneously waiting for a user's interaction to report the changes to the server-side with new information blocks *c(k)*.

Regarding the software solutions for the server-side, there are many options for programming the *real-time control loop* (Matlab/Simulink, C++, Scicos, etc.). In this chapter, we propose a working scheme for LabVIEW developers. The set of tasks that should be executed in the Lab-

Figure 2. Stream of information between client and server



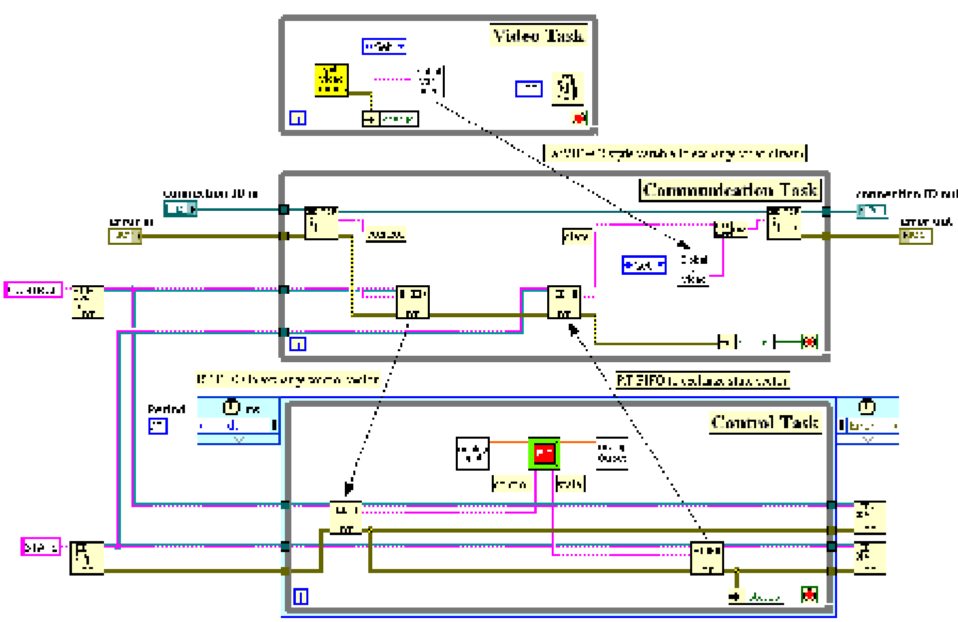
VIEW server to enable the remote experimentation are described below:

- **Control task:** This loop involves the execution of three sub-tasks: 1) recover control parameters from the communication task, 2) acquire data and closed-loop control, and 3) transmit the system state to the communication task.
- **Video task:** This loop involves the execution of two sub-tasks: 1) acquire images from the video camera and 2) transmit images to the communication task.
- **Communication task:** This loop involves the execution of three sub-tasks: 1) receive control data from clients and write the data to the control task, 2) read the system state from the control task and the images from the video task, and 3) link the state and images and then send them to the client.

Figure 3 shows a LabVIEW block diagram that corresponds to this communication architecture. The three loops in the diagram run concurrently to perform the main tasks: *control*, *communication*, and *video acquisition*.

The control task is a *time-critical* activity running at a sampling period of 20 ms with a priority greater than the other two threads. The *Analog Input Block* reads the analog input signal from the sensor, its output is compared to the setpoint input of the *PID Block,* and the result is fed into the *Analog Output Block*. Then the resulting value is sent to the actuator, which completes the control task. The data structure composed of the setpoint value, the PID control parameters, the command to the actuator, and other variables is known as the *control vector*. This vector is sent from the communication task to the control task through *RT FIFO* queue blocks (RT FIFO queues act as a fixed-size queue so that the writing of data to an RT FIFO does not overwrite previous elements). These variables are produced when users interact with the client interface. The data array formed by the values sent to the actuator, the measurement from the sensor, the current time, and other variables are known as the state vector, and these values are transferred from the control

Figure 3. Three loops running concurrently in the LabVIEW server



task to the communication task through RT FIFO queue blocks.

The video task is a *non-time-critical* activity because the loss of some video frames is generally acceptable for the user. For most applications, sending five images per second is enough to obtain adequate visual feedback of the remote system (Salzmann et al. 1999).

The communication task concatenates the current measurements (state vector) and the video frame in a new vector. This resulting vector is sent to the client using a *TCP Write Block*. In parallel, the control vector is received through the *TCP Read Block* from the clients and is passed to the control task through *RT FIFO* queues. The TCP protocol is used in both implementations because it guarantees packet delivery and bandwidth adaptation, although there is the cost of extra transmission delays (Lim 2006). A possible alternative would be the UDP protocol, which provides better control of the transmission delay. However, UDP does not have a guaranteed packet delivery mechanism or a bandwidth adaptation mechanism. Thus, the designer is responsible for implementing these features.

Once the server-side is completed, the EJS application on the client side must be modified to exchange information with the LabVIEW server. In other words, the virtual lab must be transformed into a remote lab by receiving data from the real system instead of the simulated one. The steps to allow a virtual lab connect to the server architecture will be explained in the next section.

First, a set of Java methods were programmed in EJS to control the connection with the LabVIEW server. Table 1 shows an example of the implementation of the methods connect(), disconnect(), sender(), and receiver().

Specifically, the upper part of Table 1 shows the excerpts of Java code that are used for establishing and releasing the connection with a server computer. TCP sockets are used to access the network layer. In Java, the socket programming creates an object and generates calls to the methods for the object. On the left part, the establishment of the connection is carried out in Line 4. To create a socket object, the *domain name* (or IP address) and the *service port* in the remote server are needed. Then, in Lines 5 and 6, the input/output stream buffers are created. These buffers act as FIFO storing queues whose filling and emptying depend on possible delays in network communication. The disconnection from the server is made by invoking the close() method in the socket and the input/output stream buffers (Lines 5, 6, 7 - right upper part of Table 1).

Conversely, the receiver() and sender() methods should be launched on independent Java threads when the connect() method is started. The sender() method is used to report changes in the user *view* that affect the operation of the remote equipment (for example, a change in a controller parameter). The receiver() method recovers the incoming data sent from the LabVIEW server. As shown in both pieces of code, the format of variables for the exchange must be defined. For the receiver() method, the *current time* (*t*), *liquid level* (*h*), and *input flow* in automatic mode (*qautomatic*) are received. These values constitute the states of the system (measurements). In the case of the sender() method, the *control mode* (*m/a*), the *input flow* in manual mode (*qman*), the *PID parameters* (*Kp*, *Ti* and *Td*), and the *setpoint* value (*ref*) are sent to the server-side when any user interaction is detected. These data are rendered to the client with an EJS *view*. The interface contains the same graphical elements of the virtual lab, but now the dynamic behavior of the elements is updated using the measurements obtained from the server.

Once the methods have been added to the EJS client, the programming logic that discriminates between working in a simulation or a remote mode has to be created. This logic depends on whether updating the variables in the hybrid lab is carried out based on the evolution of the mathematical model (virtual lab) or on real measurements that are obtained from the server when the remote working mode is active (remote lab).

Table 1. Excerpt of Java code to communicate with the LabVIEW server from the EJS client

Connect with the server

```
public boolean connect(){
connected = false;
try{
javaSocket = new Socket(“onetank.dia.uned.es”, 2055);
in = new DataInputStream(javaSocket.getInputStream());
out = new DataOutputStream(javaSocket.getOutputStream());
if (javaSocket != null) { // If connected ?
connected = true; // connection is ok...
_play(); // executing evolution
}
}catch (java.net.IOException io) {
System.out.println(“Problems connecting to host.”);
}
return connected;
}
```

Disconnect from the server

```
public void disconnect(){
if (connected) {
if (javaSocket != null){
try {
in.close();// close input stream
out.close();// close output stream
javaSocket.close();// close connection
javaSocket = null;
in = null;
out = null;
connected = false;
}catch (java.io.IOException e){
System.out.println(“Close socket error.”);
}
}
}
}
```

Receive data from the server

```
public void receiver(){
if (connected) {
try {
time = in.readFloat();//read time from server
level = in.readFloat();// read level from server
qautomatic = in.readFloat();// read input flow from server
}catch (java.io.IOException e) {
System.out.println(“Error receiving data.”);
}
}
}
```

Send data to the server

```
public void sender(){
if (connected) {
try {
out.writeBoolean(m/a);//write control mode
out.writeFloat(qman);//write input flow in manual
out.writeFloat(Kp);//write proportional gain
out.writeFloat(Ti);//write integral gain
out.writeFloat(Td);//write derivative gain
out.writeFloat(ref);//write setpoint
out.flush();// flush data to client
}catch (java.io.IOException e) {
System.out.println(“Error sending data.”);
}
}
}
```

Integrating Hybrid Laboratories into a Learning Management System

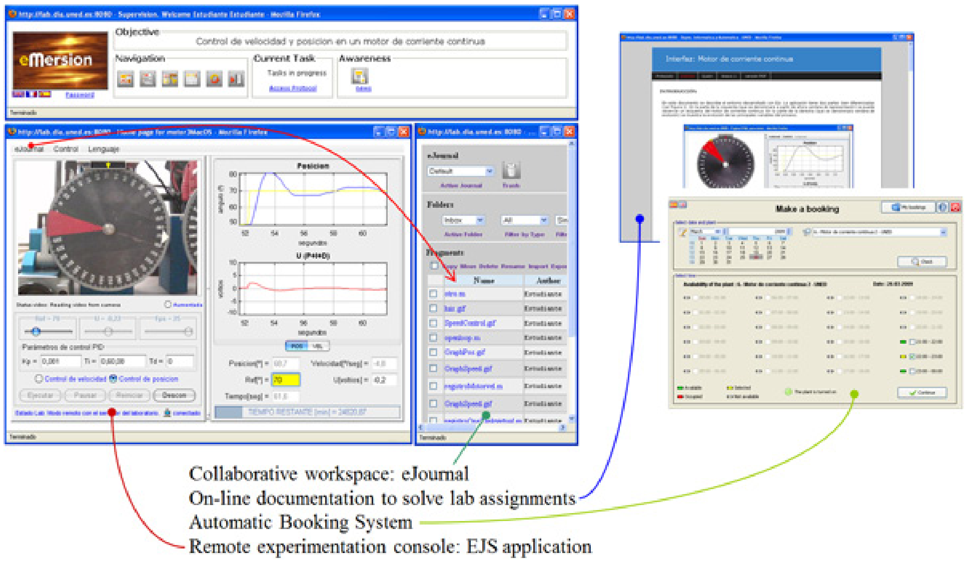
Remote and virtual control laboratories do not provide all the necessary resources to teach students in a distributed scenario. This section describes the Web infrastructure used to support the learning process of students in a distributed scenario. eMersion (Gillet et al. 2005) is the LMS tool we chose for publishing the virtual and remote laboratories on the Internet. This environment was implemented based on emulating the social behavior of the interactions and collaborations that exist in an f2f laboratory.

eMersion Description

Figure 4 shows a complete view of the experimentation environment for eMersion during a practical session with a DC motor system in remote mode. From a structural point of view, the environment is composed of five independent Web applications: *navigation bar*, *eJournal*, *experimentation console*, *online information*, and *external applications*.

The *navigation bar* provides access to the other Web resources for the environment. From the link labeled “Access Protocol,” users can obtain a complete user’s guide for the environment.

Figure 4. Learning Management System to publish web-based labs



The *eJournal* resource provides a shared workspace for users to communicate and collaborate during the learning process. The *eJournal* allows students to save, retrieve, and share their experimental results and documents. Furthermore, the presentation of results and discussions with teaching staff can be performed using the options that are provided. The users can also organize the information collected during the experimental sessions and through online repositories. Work tracking and awareness can be implemented based on this information.

The *experimentation console* corresponds to the EJS interfaces in which students carry out their experimental activities. These interfaces can interchange data with the *eJournal* space (see Figure 4). Thus, students can use the results they obtained (through images of the system's evolution or data registers) in the experimentation sessions to prepare their reports for the final assessment.

Online information is a collection of HTML pages and PDF files that allow students to visualize all the documentation necessary to solve the laboratory assignments.

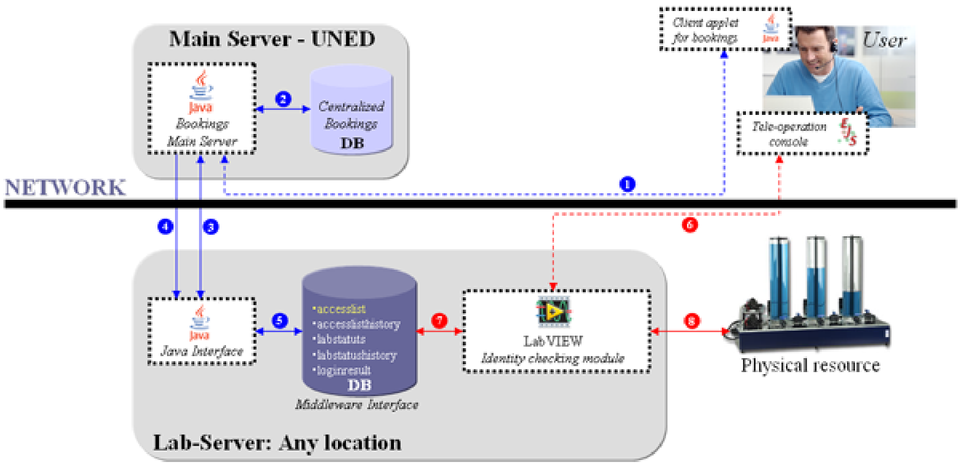
Finally, eMersion offers an ability to integrate external Web applications. In this context, the following subsection describes the *automatic bookings* and *authentication system* developed to organize students' access to the physical resources. This application was successfully integrated into eMersion.

A Flexible Scheme for Authentication and Booking of Physical Resources

A flexible scheme to let students book the use of a physical resource located in the laboratory was added to the LMS. Essentially, students can fill out a reservations database automatically from the client through a Web interface. The system includes three main modules.

For the first module, a Java applet was developed to perform new bookings on the client-side (see *Client applet for bookings* in Figure 5). For the second module, a centralized server application was also developed in Java to manage reservations, synchronism, and communications between the client applet for bookings and the Lab-Server (see *Bookings Main Server* in Figure 5). Finally, an

Figure 5. A flexible scheme for bookings and the authentication process



additional Java module located in the Lab-Server was developed (see *Java Interface* in Figure 5). This module informs the Bookings Main Server of the current state of the Lab-Server and other parameters that the central server requires.

The full process for booking a physical resource in the laboratory is divided into two stages, which are described below (see Figure 5):

Reservation Phase: A description of the states flow during a reservation occurs as follows:

- The Applet for bookings requests a new reservation (step 1)
- The Bookings Main Server takes the request and saves it in a local database (DB) (step 2)
- The Bookings Main Server asks the Java Interface for its time zone (step 3)
- The Java Interface provides its time zone to the Bookings Main Server (step 3)
- The Bookings Main Server calculates the time lag and amends the timeslot
- The Bookings Main Server reports the new register to the Java Interface (step 4)
- The Java Interface receives the register and inserts it in the Lab-Server DB (step 5)
- The Bookings Main Server tells the client that the new reservation has been made (step 1)

Authentication Phase: A description of the states flow during authentication occurs as follows:

- The Applet for experimentation starts the process by sending user credentials (step 6)
- The Identity Checking Module receives the keys and checks whether the user exists in the local DB. If the user exists, it then checks whether the connection attempt is between the start-time and the end-time of the timeslot reserved (step 7)
- The Identity Checking Module sends the result of the checking to the Applet for experimentation (step 6)
- If the checking result is acceptable, then the Applet for experimentation receives free access to the Target Plant (steps 6 and 8)

The *Applet for bookings* in which students schedule their reservations for any experiment is shown in Figure 4. The process to make a booking requires that when the student requests a reservation, the response of the bookings system must indicate the date and time assigned for the student to use the remote plant.

Other booking and authentication systems with similar features can be found in the literature. One of the most relevant is the booking and authentication mechanism for the iLabs Shared System (http://icampus.mit.edu/ilabs/), which was developed at the Massachusetts Institute of Technology (MIT). This program employs a middle tier Service Broker that manages the interaction between users and Lab-Servers. In this architecture, all of the reservations are hosted in the service broker and the users must pass through it each time they want to work with the real plants. Unlike the iLabs system, our system offers some additional features. When a user performs a new booking, the reservation is hosted and managed in a middle tier, which is similar to the iLabs Shared System, but the reservation is also stored in a simple database located in each Lab-Server. Thus, the post authentication process is carried out directly with the Lab-Server by passing this middle tier. Another advantage of this architecture is that in case a Lab-Server with valid bookings is damaged, these bookings could be retrieved later by the Lab-Server from the Central Server. Finally, the administrators of a Lab-Server could also manage the bookings locally. As such, if there were problems with the central servers, bookings could be made manually.

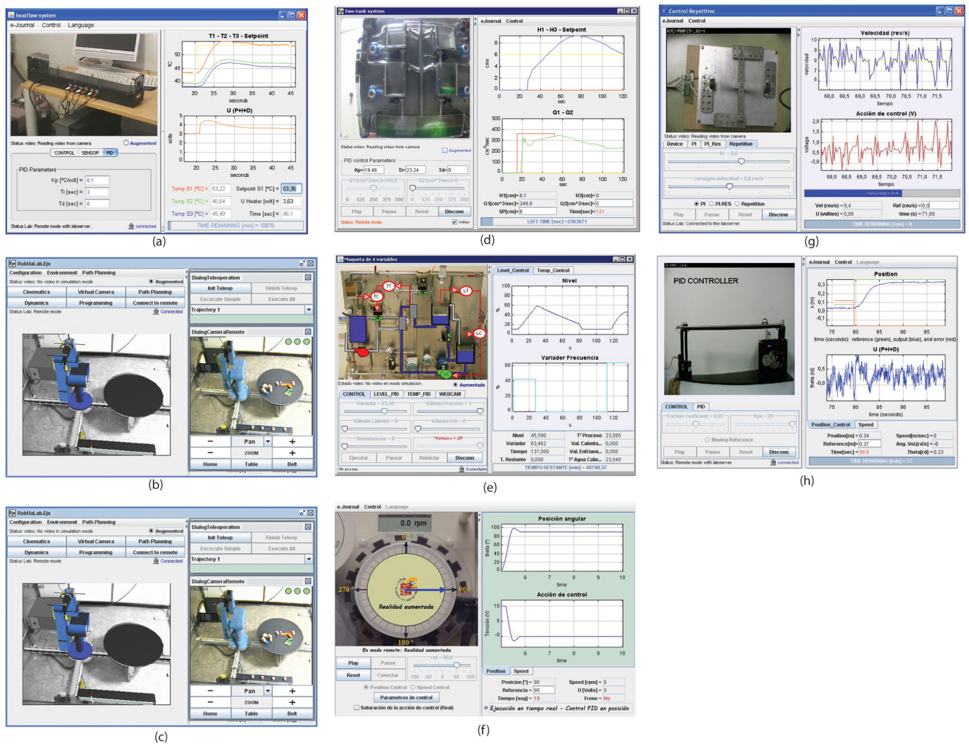
AUTOMATL@BS NETWORK

The AutomatL@bs network (http://lab.dia.uned.es/automatlab) is a consortium of seven Spanish universities that decided to expand their efforts in the use of virtual and remote laboratories for engineering education to a national level. The universities taking part in this project are UNED, the University of Almería (UAL), the University of Alicante (UA), Polytechnic University of Valencia (UPV), Polytechnic University of Catalonia (UPC), Miguel Hernández University (UMH), and the University of León (UNILEON). The main challenge of this work has been to manage and coordinate the integration of the *hardware*, *software*, and *human* resources in a Web-based experimentation environment hosted by the Department of Computer Science and Automatic Control of UNED in Madrid. The main aims of this project were

- Enabling students to access practical experiments that are not available at their universities.
- Increasing the quality and robustness of the network of virtual and remote laboratories for a higher number of students and teachers with different teaching concerns.

The Web-based laboratories were offered to a total of 112 master's degree candidates at engineering schools in the consortium. Figure 6 shows the GUIs of the virtual and remote laboratories for each participant university. Each GUI has the same arrangement of graphical elements, which provides a uniform structure. Also, the Web-based laboratories of the AutomatL@bs network were documented following the same guidelines and criteria. In general, the documentation defines a set of tasks or activities that students should carry out so they can be evaluated effectively by the teaching staff. This sequence of activities was divided into two phases: *PRE-Lab activities* and *Lab activities*. PRE-Labs are based on the use of the experimentation console in the simulation mode. Thus, the teaching team could be assured that each student had prior knowledge of the system before using it for an actual experiment. Lab tasks are based on the use of the experimentation console in the remote mode. Remote access to the system is allowed by the teaching staff once students finish their PRE-lab work in simulation mode, and the work is considered satisfactory.

Figure 6. The available remote systems in AutomatL@bs



Description of the Pedagogical Scenario

Students from each university worked on three of the nine available remote systems (three from UNED and six from other universities) offered by the AutomatL@bs project (one lab from their university and two labs from other locations). Then students were required to use the system to learn how to operate the interfaces. Finally, after several sessions, the students could complement their work with the Web-based AutomatL@bs experimentation system at their convenience through an outside Internet connection. UNED students were first offered the chance to access the systems that were available at UNED (a *servo-motor system*, a *heat-flow system*, and a *three-tank system*). Later on, they could complete their work remotely through the Internet. During these experimental sessions, students were able to save their data measurements and parameters for writing their final reports. The students placed their reports in the *eJournal* space for evaluation. Teaching assistants from each university were in charge of evaluation.

Outcomes

To obtain feedback regarding the use of the system, the students were required to complete evaluation questionnaires. We designed questions based on the guidelines of Ma and Nickerson (2006) to evaluate the infrastructure and the technical quality of the system as well as the educational value and the experiences of students. Some of the more relevant questions are described below:

Technical questions:

1. How would you describe the quality of the virtual laboratories?
2. How would you describe the quality of the remote laboratories?

3. Have you experienced hardware or software problems?
4. Did you appreciate the uniform structure of the client interfaces?
5. How was the navigation experience for the global system options?

Educational value questions:

1. How would you evaluate the quality of your learning with Web-based laboratories compared to traditional methods?
2. How would you describe your learning speed using remote and virtual labs compared to traditional methods?
3. In general terms, are you satisfied with the usability of the system?
4. What were the most important learning resources when you were learning to use the system?
5. How would you evaluate the level of difficulty of using the system?

Although this assessment was not an exhaustive evaluation, it provided initial information on what was necessary, or unnecessary, to include in this methodology for future engineering courses. The outcomes obtained from this survey are summarized in Table 2.

Sub-scale Number 1 provided the first general view concerning whether students felt satisfied with this new method for performing their practical experiments. The results showed that 19% and 69% of students strongly agreed and agreed, respectively, that they were satisfied with the system. Other questions about the advantages of using remote experiments in the educational process were also reported. The results showed that the use of new technologies, especially the Internet, encouraged students to conduct most of their practical exercises using this resource. *Sub-scale Numbers 2 and 3* show comparative information about learning with the new technological methods compared to traditional methods. In cases where students reported dissatisfaction (9%), the primary reason was that they were not able to work directly with the laboratory equipment. A way to solve this problem could be to apply an educational methodology based on *blended learning*. First, a face-to-face class where students could interact and experiment in situ with the real plant would be held. The students would then be allowed access to the experimental environment remotely to complete their practical exercises. Regarding the quality of the hybrid laboratories (*Sub-scale Numbers 4 and 5*), most students positively evaluated their development in terms of user functionality. Any negative results might

Table 2. Summary of the survey outcomes

Sub-scale	A%	B%	C%	D%	E%
1. Satisfaction degree	19	69	7	5	0
2. Learning compared to traditional methods	15	51	25	8	1
3. Facility of using the system	19	62	11	8	0
4. Quality of virtual labs	33	48	15	4	0
5. Quality of remote labs	25	38	25	10	2
6. Most important learning resource	18	44	27	11	0

1. A: Strongly Agree B: Agree C: Neutral D: Disagree E: Strongly Disagree
2. A: Much better B: Better C: Equal D: less E: Much less
3. A: Strongly Agree B: Agree C: Neutral D: Disagree E: Strongly Disagree
4. A: Very good B: Good C: Acceptable D: Bad E: Very Bad
5. A: Very good B: Good C: Acceptable D: Bad E: Very Bad
6. A: Documentation B: Questions to teacher C: Simulation D: Connection to plant E: Others

have been a consequence of the quality of the Internet connection speed because slow speeds lead to delays. Some of the students performed their experiments using old dialup connections (56 kbps), so the exchange of data with some processes was not fast enough and caused the user interfaces to update slowly. The experiments were also tested with low-speed ADSL lines (512/128 kbps), and the results were satisfactory. Finally, *Sub-scale Number 6* shows how queries of the teaching staff and the documentation of the practical exercises were essential resources for positive student performance.

CONCLUSION

Virtual and remote experimentation for engineering education can be considered a mature technology. However, the process of transforming a classic control experiment into an interactive Web-based laboratory is not an easy task. This essay provides a systematic approach for developing prototypes of remote laboratories from a pedagogical perspective using three tools: EJS, LabVIEW, and eMersion. This approach incorporates the development of online experimentation environments and provides an effective scheme to switch between the *simulation* and *teleoperation* of real systems.

The *AutomatL@bs project* has yielded benefits to the universities that have participated in the project over the last three academic years. The results from the previous evaluation allowed us to debug the system and identify necessary improvements in the framework. First, the number and variety of available experiments will be increased by enrolling new universities in the AutomatL@bs project (with special interest in universities from South America). To cope with this challenge, other LMS, such as Moodle or Sakai, are currently being evaluated. Second, the applets and all the materials are being adapted to the SCORM standards to simplify porting to another LMS. Additionally, the applets of the simulated physical processes are being integrated into the ComPADRE digital library (http://www.compadre.org) to gain visibility. These changes could help integrate and deploy our project in other institutions. We will also attempt to let students carry out their practical experiments using other devices (such as mobile phones and PDAs) and user interfaces (including e-mail, Web forms, and HTML/JavaScript thin interfaces).

REFERENCES

Astrom, K. J. (2006, June). *Challenges in control education*. Paper presented at the 7th IFAC Symposium on Advances in Control Education (ACE), Madrid, Spain.

Bourne, J., Harris, D., & Mayadas, F. (2005). Online engineering education: Learning anywhere, anytime. *International Journal of Engineering Education*, *91*(1), 131–146.

Brito, N., Ribeiro, P., Soares, F., Monteiro, C., Carvalho, V., & Vasconcelos, R. (2009, November). *A remote system for water tank level monitoring and control - a collaborative case-study*. Paper presented at the 3rd IEEE International Conference on e-Learning in Industrial Electronic (ICELIE), Porto, Portugal.

Callaghan, M. J., Harkin, J., McGinnity, T. M., & Maguire, L. P. (2006). Client-server architecture for remote experimentation for embedded systems. *International Journal of Online Engineering*, *2*(4), 8–17.

Casini, M., Prattichizzo, D., & Vicino, A. (2004). The automatic control telelab. A web-based technology for distance learning. *IEEE Control Systems Magazine*, *24*(3), 36–44. doi:10.1109/MCS.2004.1299531

Christian, W., & Esquembre, F. (2007). Modeling physics with Easy Java simulations. *The Physics Teacher*, *45*(10), 475–480. doi:10.1119/1.2798358

Dormido, S. (2004). Control learning: Present and future. *Annual Reviews in Control*, *28*(1), 115–136. doi:10.1016/j.arcontrol.2003.12.002

Dormido, S., Canto, S. D., Canto, R. D., & Sánchez, J. (2005). The role of interactivity in control learning. *International Journal of Engineering Education: Special Issue on Control Engineering Education*, *21*(6), 1122–1133.

Easy Java. (2010). *EJS wiki homepage*. Retrieved November 10, 2010, from http://www.um.es/fem/EjsWiki/

Eikaas, T. I., Foss, B. A., Solbjorg, O. K., & Bjolseth, T. (2006). Game-based dynamic simulations supporting technical education and training. *International Journal of Online Engineering*, *2*(2), 1–7.

EMERSION. (2010). *eMersion project homepage*. Retrieved November 10, 2010, from http://lawww.epfl.ch/page28147.html

Esquembre, F. (2005). *Creación de simulaciones interactivas en Java*. Madrid, Spain: Pearson, Prentice Hall.

Gillet, D., El Helou, S., Marie, J., & Rosamund, S. (2009, September). *Science 2.0: Supporting a doctoral community of practice in technology enhanced learning using social software*. Paper presented at the 4th European Conference on Technology Enhanced Learning (EC-TEL), Nice, France.

Gillet, D., Nguyen, A. V., & Rekik, Y. (2005). Collaborative web-based experimentation in flexible engineering education. *IEEE Transactions on Education*, *48*(4), 696–704. doi:10.1109/TE.2005.852592

LABVIEW. (2010). *NI LabVIEW homepage*. Retrieved November 10, 2010, from http://www.ni.com/labview/

Lim, D. (2006). A laboratory course in real-time software for the control of dynamic systems. *IEEE Transactions on Education*, *49*(3), 346–354. doi:10.1109/TE.2006.879243

Ma, J., & Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Computing Surveys*, *38*(3), 1–24.

Nguyen, A. V. (2007, July). *Activity theoretical analysis and design model for web-based experimentation*. Paper presented at the 12th International Conference on Human-Computer Interaction, Beijing, China.

Oppenheim, A., Willsky, A., & Hamid, S. (1996). *Signals and systems* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Rosen, M. A. (2007). Future trends in engineering education. In Aung, W. (Eds.), *Innovations 2007: World innovations in engineering education and research* (pp. 1–11). Arlington, VA: International Network for Engineering Education and Research/Begell House Publishing.

Salzmann, C., & Gillet, D. (2002, July). *Real-time interaction over the Internet*. Paper presented at the 15th IFAC World Congress, Barcelona, Spain.

Salzmann, C., & Gillet, D. (2008). From online experiments to smart devices. *International Journal of Online Engineering*, *4*(S1), 50–54.

Salzmann, C., Gillet, D., & Huguenin, P. (1999). Introduction to real-time control using LabVIEW with an application to distance learning. *International Journal of Engineering Education*, *16*(3), 255–272.

Valera, A., Diez, J. L., Vallés, M., & Albertos, P. (2005). Virtual and remote control laboratory development. *IEEE Control Systems Magazine*, *25*(1), 35–39. doi:10.1109/MCS.2005.1388798

Vargas, H., Sánchez, J., Duro, N., Dormido, R., Dormido-Canto, S., & Farias, G. (2008). A systematic two-layer approach to develop web-based experimentation environments for control engineering education. *Intelligent Automation and Soft Computing*, *14*(4), 505–524.

Williams, R. (2007). Flexible learning for engineering. In Aung, W. (Eds.), *Innovations 2007: World innovations in engineering education and research* (pp. 279–290). Arlington, VA: International Network for Engineering Education and Research/Begell House Publishing.

Zutin, D. G., Auer, M. E., Bocanegra, J. F., López, E. R., Martins, A. C. B., Ortega, J. A., & Pester, A. (2008). TCP/IP communication between server and client in multi user remote labs applications. *International Journal of Online Engineering*, *4*(3), 42–45.

ADDITIONAL READING

Abdulwahed, M., & Nagy, Z. K. (2009). Applying Kolbs experiential learning on laboratory education, case study. *Journal of Engineering Education*, *98*(3), 283–294.

Aliane, N., Pastor, R., & Mariscal, G. (2010). Limitations of remote laboratories in control engineering education. *International Journal of Online Engineering*, *6*(1), 31–33.

Christian, W., Esquembre, F., & Mason, B. (2009, September). *Easy Java Simulations and the ComPADRE library*. Paper presented at the 14th International Workshop on Multimedia in Physics Teaching and Learning (MPTL14), Udine, Italy.

de la Torre, L., Sánchez, J., & Dormido, S. (2009, September). *The Fisl@bs portal: A network of virtual and remote laboratories for physics education*. Paper presented at the 14th International Workshop on Multimedia in Physics Teaching and Learning (MPTL14), Udine, Italy.

Dormido, R., Vargas, H., Duro, N., Sánchez, J., Dormido-Canto, S., & Farias, G. (2008). Development of a web-based control laboratory for automation technicians: The three-tank system. *IEEE Transactions on Education*, *51*(1), 35–44. doi:10.1109/TE.2007.893356

Duro, N., Dormido, R., Vargas, H., Dormido-Canto, S., Sánchez, J., Farias, G., & Dormido, S. (2008). An integrated virtual and remote control lab: The three-tank system as a case study. *Computing in Science & Engineering*, *10*(4), 50–59. doi:10.1109/MCSE.2008.89

Fakas, G. J., Nguyen, A. V., & Gillet, D. (2005). The electronic laboratory journal: A collaborative and cooperative learning environment for web-based experimentation. *Computer Supported Cooperative Work*, *14*(3), 189–216. doi:10.1007/s10606-005-3272-3

Gillet, D., Nguyen, A. V., & Rekik, Y. (2005). Collaborative web-based experimentation in flexible engineering education. *IEEE Transactions on Education*, *48*(4), 696–704. doi:10.1109/TE.2005.852592

Gomes, L., & Bogosyan, S. (2007). Special section on e-learning and remote laboratories within engineering education - first part. *IEEE Transactions on Industrial Electronics*, *54*(6), 3054–3056. doi:10.1109/TIE.2007.907007

Guzmán, J. L., Vargas, H., Sánchez, J., Berenguel, M., Dormido, S., & Rodríguez, F. (2007). Education research in engineering studies: Interactivity, virtual and remote labs. In Morales, A. V. (Ed.), *Distance Education Issues and Challenges* (pp. 131–167). Hauppauge, NY: Nova Science Publishers.

ILOUGH-LAB. (2010). *The Ilough-lab. A Process Control Lab in the Chemical Engineering Lab at Loughborough University, UK*. Retrieved November 10, 2010, from http://www.ilough-lab.com.

ISILAB. (2010). *ISILab Internet Shared Instrumentation Laboratory University of Genoa*. Retrieved November 10, 2010, from http://isilab.dibe.unige.it.

Jara, C., Esquembre, F., Candelas, F., Torres, F., & Dormido, S. (2009, October). *New features of Easy Java Simulations for 3D modelling*. Paper presented at the 8th IFAC Symposium on Advances in Control Education (ACE09), Kumamoto, Japan.

LABSHARE. (2010). *LabShare. University of Technology, Sydney*. Retrieved November 10, 2010, from http://www.labshare.edu.au.

Lareki, A., Martínez, J., & Amenabar, N. (2010). Towards an efficient training of university faculty on ICTs. *Computers & Education*, *54*(2), 491–497. doi:10.1016/j.compedu.2009.08.032

Martín, C., Urquía, A., & Dormido, S. (2007). Object-oriented modelling of virtual-laboratories for control education. In Tzafestas, S. G. (Ed.), *Web-based Control and Robotics Education* (pp. 103–125). Springer Verlag. doi:10.1007/978-90-481-2505-0_5

Nguyen, A. V., Rekik, Y., & Gillet, D. (2006). Iterative Design and Evaluation of a Web-Based Experimentation Environment. In Lambropoulus, N., & Zaphiris, P. (Eds.), *User-Centered Design of Online Learning Communities* (pp. 286–313). Idea Group. doi:10.4018/978-1-59904-358-6.ch013

NUS. (2010). *Internet Remote Experimentation. National University of Singapore*. Retrieved November 10, 2010, from http://vlab.ee.nus.edu.sg/~vlab/.

Restivo, M.T., & Silva, M.G. (2009). Portuguese Universities Sharing Remote Laboratories. *International Journal of Online Engineering, 5*(special issue IRF'09), 16-19.

Tan, K. K., Wang, K. N., & Tan, K. C. (2005). Internet-based resources sharing and leasing system for control engineering research and education. *International Journal of Engineering Education*, *21*(6), 1031–1038.

TELELAB. (2010). *Automatic Control TeleLab. University of Siena*. Retrieved November 10, 2010, from http://act.dii.unisi.it.

Vargas, H., Salzmann, Ch., Gillet, D., & Dormido, S. (2009, October). *Remote experimentation mashup*. Paper presented at the 8th IFAC Symposium on Advances in Control Education (ACE09), Kumamoto, Japan.

Vargas, H., Sánchez, J., Salzmann, Ch., Esquembre, F., Gillet, D., & Dormido, S. (2009). Web-enabled remote scientific environments. *Computing in Science & Engineering*, *11*(3), 34–46. doi:10.1109/MCSE.2009.61

KEY TERMS AND DEFINITIONS

Control Engineering: This form of engineering is an engineering discipline that applies control theory to design systems with predictable behaviors. The practice of control engineering uses sensors to measure the output performance of the device that is being controlled (e.g., a process or a vehicle), and those measurements can be used to provide feedback to the input actuators that can make corrections toward the desired performance.

Controller: This is a device that monitors and affects the operational conditions of a given dynamical system. The operational conditions are typically referred to as output variables in the system that can be affected by adjusting certain input variables.

Hybrid Laboratories: A Web-based laboratory where students can work with a simulation of a dynamic system (virtual lab) or on the real counterpart (remote lab).

LabVIEW (Laboratory Virtual Instrumentation Engineering Workbench): LabVIEW is a graphical programming environment from National Instruments used to develop sophisticated measurements, tests, and control systems with intuitive graphical icons and wires that resemble a flowchart.

Model–View–Controller (MVC): This is a software architecture that is currently an architectural pattern used in software engineering.

PID Controller: A proportional–integral–derivative controller is a generic control loop feedback mechanism (controller) widely used in industrial control systems.

Sharable Content Object Reference Model (SCORM): A collection of standards and specifications for Web-based e-learning. It defines communications between client-side content and a host system called the run-time environment, which is commonly supported by a learning management system.

This work was previously published in Internet Accessible Remote Laboratories: Scalable E-Learning Tools for Engineering and Science Disciplines, edited by Abul K.M. Azad, Michael E. Auer and V. Judson Harward, pp. 206-225, copyright 2012 by Engineering Science Reference (an imprint of IGI Global).

Chapter 40
An Estimation of Distribution Algorithm for Part Cell Formation Problem

Saber Ibrahim
University of Sfax, Tunisia

Bassem Jarboui
University of Sfax, Tunisia

Abdelwaheb Rebaï
University of Sfax, Tunisia

ABSTRACT

The aim of this chapter is to propose a new heuristic for Machine Part Cell Formation problem. The Machine Part Cell Formation problem is the important step in the design of a Cellular Manufacturing system. The objective is to identify part families and machine groups and consequently to form manufacturing cells with respect to minimizing the number of exceptional elements and maximizing the grouping efficacy. The proposed algorithm is based on a hybrid algorithm that combines a Variable Neighborhood Search heuristic with the Estimation of Distribution Algorithm. Computational results are presented and show that this approach is competitive and even outperforms existing solution procedures proposed in the literature.

INTRODUCTION

The principle objective of Group Technology is to reduce the intercellular flow of parts and to provide an efficient grouping of machines into cells. The main contribution in this chapter is to develop an efficient clustering heuristic based on evolutionary algorithms and to apply the proposed heuristic for Machine Part Cell Formation Problem which includes the configuration and capacity management of manufacturing cells. We propose to apply a novel population based evolutionary algorithm called Estimation of Distribution Algorithm in order to form part families and machine cells simultaneously.

DOI: 10.4018/978-1-4666-1945-6.ch040

The objective of the proposed heuristic is to minimize exceptional elements and to maximize the goodness of clustering and thus the minimization of intercellular movements. In order to guarantee the diversification of solutions, we added an efficient technique of local search called Variable Neighborhood Search at the improvement phase of the algorithm. Many researchers have combined local search with evolutionary algorithms to solve this problem. However, they did not apply yet the Estimation of Distribution Algorithm for the general Group Technology problem. Furthermore, we have used a modified structure of the probabilistic model within the proposed algorithm.

In order to quantify the goodness of the obtained solutions, we present two evaluation criteria namely the percentage of exceptional elements and the grouping efficacy. A comparative study was elaborated with the most known evolutionary algorithms as well as the well known clustering methods.

LITERATURE REVIEW

A wide body of publications has appeared on the subject of Group Technology (GT) and Cellular Manufacturing Systems (CMS). The history of approaches that tried to solve this problem began with the classification and coding schemes. Several authors have proposed various ways trying to classify the methods of Cell Formation Problem. It includes descriptive methods, cluster analysis procedures, graph partitioning approaches, mathematical programming approaches, artificial intelligence approaches and other analytical methods.

Burbidge (1963) was the first who developed a descriptive method for identifying part families and machine groups simultaneously. In his work "Production Flow Analysis" (PFA). Burbidge has proposed an evaluative technique inspired from an analysis of the information given in route cards to find a total division into groups, without any need to buy additional machine tools.

Then, researchers applied array based clustering techniques which used a binary matrix *A* called "Part Machine Incidence Matrix" (PMIM) as input data. Given *i* and *j* the indexes of parts and machines respectively, an entry of 1 (a_{ij}) means that the part *i* is executed by the machine *j* whereas an entry of 0 indicates that it does not. The objective of the array based techniques is to find a block diagonal structure of the initial PMIM by rearranging the order of both rows and columns. Thus, the allocation of machines to cells and the parts to the corresponding families is trivial. McCornick et al. (1972) were the first who applied this type of procedure to the CFP. They developed the Bond Energy Analysis (BEA) which seeks to identify and display natural variable groups and clusters that occur in complex data arrays. Besides, their algorithm seeks to uncover and display the associations and interrelations of these groups with one another. King (1980) developed the Rank Order Clustering (ROC). In ROC algorithm, binary weights are assigned to each row and column of the PMIM. Then, the process tries to gather machines and parts by organizing columns and rows according to a decreasing order of their weights. Chan and Milner (1981) developed the Direct Clustering Algorithm (DCA) in order to form component families and machine groups by restructuring the machine component matrix progressively. A systematic procedure is used instead of relying on intuition in determining what row and column rearrangements are required to achieve the desired result. King & Nakornchai (1982) improved the ROC algorithm by applying a quicker sorting procedure which locates rows or columns having an entry of 1 to the head of the matrix. Chandrasekharan & Rajagopalan (1986a) proposed a modified ROC called MODROC, which takes the formed cells by the ROC algorithm and applies a hierarchical clustering procedure to them. Later, other array based clustering techniques are proposed namely the Occupancy Value method of Khator & Irani (1987), the Cluster Identification Algorithm (CIA) of Kusiak & Chow (1987) and the Hamiltonian Path Heuristic of Askin et al. (1991).

McAuley (1972) was the first who suggested similarity coefficient to clustering problems. He applied the Single Linkage procedure to the CF problem and used the coefficient of Jaccard which is defined for any pair of machines as the ratio of the number of parts that visit both machines to the number of parts that visit at least one of these machines. Then, some other clustering techniques are developed namely Single Linkage Clustering (SLC), Complete Linkage Clustering (CLC), Average linkage Clustering (ALC) and Linear Cell Clustering (LCC). Kusiak (1987) proposed a linear integer programming model maximizing the sum of similarity coefficients defined between two parts

The category that is the most used in literature in recent years is heuristics and metaheuristics. Such heuristics are based essentially on Artificial Intelligence approaches including Genetic Algorithms (GA), Simulated Annealing (SA), Tabu Search (TS), Evolutionnary Algorithms (EA), neural network and fuzzy mathematics. In what follows we present some research papers that used this type of heuristics for designing CM systems.

Boctor (1991) developed the SA approach to deal with large-scale problems. Sofianopoulos (1997) proposed a linear integer formulation for CF problem and employed the SA procedure to improve the solution quality taking as objective the minimization of inter-cellular flow between cells. Caux et al. (2000) proposed an approach combining the SA method for the CF problem and a branch-and-bound method for the routing selection. Lozano et al. (1999) presented a Tabu Search algorithm that systematically explores feasible machine cells configurations determining the corresponding part families using a linear network flow model. They used a weighted sum of intra-cell voids and inter-cellular moves to evaluate the quality of the solutions. Solimanpur et al. (2003) developed an Ant colony optimization algorithm to solve the inter cell layout problem by modelling it as a quadratic assignment problem. Kaparthi et al. (1993) proposed an algorithm based on neural network for the part machine grouping problem. Xu & Wang (1989) developed two approaches of fuzzy cluster analysis namely fuzzy classification and fuzzy equivalence in order to incorporate the uncertainty in the measurement of similarities between parts. They presented also a dynamic part-family assignment procedure using the methodology of fuzzy pattern recognition to assign new parts to existing part families.

Recently many researchers have focused on the approaches based on AI for solving the part-machine grouping problem. Venugopal & Narendran (1992a) proposed a bi-criteria mathematical model with a solution procedure based on a genetic algorithm. Joines et al. (1996) presented an integer programming solved using a Genetic Algorithm to solve the CF problem. Zhao & Wu (2000) presented a genetic algorithm to solve the machine-component grouping problem with multiple objectives: minimizing costs due to inter-cell and intra-cell part movements; minimizing the total within cell load variation; and minimizing exceptional elements. Gonçalves & Resende (2002) developed a GA based method which incorporates a local search to obtain machine cells and part families. The GA is responsible for generating sets of machines cells and the mission of the local search heuristic is to construct sets of machine part families and to enhance their quality. Then, Gonçalves & Resende (2004) employed a similar algorithm to find first the initial machine cells and then to obtain final clusters by applying the local search. Mahdavi et al. (2009) presented a GA based procedure to deal with the CF problem with nonlinear terms and integer variables. Stawowy (2006) developed a non-specialized Evolutionary Strategy (ES) for CF problem. His algorithm uses a modified permutation with separators encoding scheme and unique concept of separators movements during mutation. Andrés

& Lozano (2006) applied for the first time the Particle Swarm Optimization (PSO) algorithm to solve the CF problem respecting the objective the minimization of inter-cell movements and imposing a maximum cell size.

ESTIMATION OF DISTRIBUTION ALGORITHM

It was first introduced by Mühlenbein & Paaß (1996). The Estimation of Distribution Algorithm belongs to Evolutionary Algorithms family. It adopts probabilistic models to reproduce individuals in the next generation, instead of crossover and mutation operations. This type of algorithms uses different techniques to estimate and sample the probability distribution.

The probabilistic model is represented by conditional probability distributions for each variable. This probabilistic model is estimated from the information of the selected individuals in the current generation and selects good individuals with respect to their fitness. This process is repeated until the stop criterion is met. Such a reproduction procedure allows the algorithm to search for optimal solutions efficiently. However, it considerably decreases the diversity of the genetic information in the generated population when the population size is not large enough. For this reason, the incorporation of a local search technique is encouraged in order to enhance the performance of the algorithm.

As a result, the Estimation of Distribution Algorithm can reach best solutions by predicting population movements in the search space without needing many parameters. The main steps in this procedure are shown in the following pseudo code:

Estimation of Distribution Algorithm

1. Initialize the population according to some initial distribution model.
2. Form P' individuals from the current population using a selection method.
3. Build a probability model p(x) from P' individuals using both the information extracted from the selected individuals in the current population and the previously built model.
4. Sample $p(x)$ by generating new individuals from the probability model and replace some or all individuals in the current population.
5. End the search if stop criteria are met, otherwise return to Step 2.

This method can be divided into two different classes. The first class assumes that there are no dependencies between variables of the current solution during the search. These are known as non-dependency Estimation of Distribution Algorithms: Population Based Incremental Learning (Baluja, 1994) and Univariate Marginal Distribution Algorithm (Mühlenbein & Paaß, 1996). The second class takes into account these variable dependencies: Mutual Information Maximization for Input Clustering (De Bonet et al., 1997), Bivariate Marginal Distributional Algorithm (Pelikan & Mühlenbein, 1999), Factorized Distribution Algorithm (Mühlenbein et al., 1999) and the Bayesian Optimization Algorithm (Pelikan et al., 1999a).

Generally, non-dependency algorithms are expected to have a worse modelling ability than the ones with variable dependencies (Zhang et al., 2004). But combining heuristic information or local search with non-dependency algorithms can compensate for this disadvantage.

Univariate EDAs

This category assume that each variable is independent; it means that the algorithm do not consider any interactions among variables in the solution. As a result, the probability model distribution,$p(x)$, becomes simply the product of Univariate marginal probabilities of all variables in the solution and expressed as follows:

$$p(x) = \prod_{i=1}^{1} p(x_i)$$

Due to the simplicity of the model of distribution used, the algorithms in this category are computationally inexpensive, and perform well on problems with no significant interaction among variables.

In what follows, we present the well-known works related to this category.

Population Based Incremental Learning

It was proposed by Baluja (1994). The algorithm starts with initialisation of a probability vector. In each iteration, it updates and samples the probability vector to generate new solutions. The main steps in this procedure are shown in the following pseudo code:

Population Based Incremental Learning

1. Initialise a probability vector $p=\{p_1,p_2,\ldots,p_n\}$ with 0.5 at each position. Here, each pi represents the probability of 1 for the i[th] position in the solution.
2. Generate a population P of M solutions by sampling probabilities in p.
3. Select set D from P consisting of N promising solutions.
4. Estimate univariate marginal probabilities p(xi) for each xi.
5. For each i, update pi using $p_i=p_i+\lambda(p(x_i-p_i)$
6. For each i, if mutation condition passed, mutate pi using

$p_i=p_i(1-\mu)$+randon (0 or 1)μ.

7. End the search if stop criteria are met, otherwise return to Step 2.

Univariate Marginal Distribution Algorithm

Univariate Marginal Distribution Algorithm was proposed by Muhlenbein & Paaß (1996). We note that this category can be seen as a variant of Population Based Incremental Learning when λ=1 and μ=0 Different variants of Univariate Marginal Distribution Algorithm have been proposed, and the mathematical analysis of their workflows has been carried out (Muhlenbein, 1998; Muhlenbein et al., 1999; Gonzalez et al., 2002). The main steps in this procedure are shown in the following pseudo code:

Univariate Marginal Distribution Algorithm

1. Generate a population *P* composed of *M* solutions.
2. Select a set *P'* from *P* consisting of *N* promising solutions.
3. Estimate univariate marginal probabilities *p*(*x*) from *P'* for each x_i.
4. Sample*p*(*x*)to generate *M* new individual and replace *P*.
5. End the search if stop criteria are met, otherwise return to Step 2.

Bivariate EDAs

In contrast with Univariate case, the probability model contains factors involving the conditional probability of pairs of interacting variables. This class of algorithms performs better in problems, where pair-wise interaction among variable exists.

In what follows, we present the well-known works related to this category.

Mutual Information Maximization for Input Clustering

The Mutual Information Maximization for input clustering uses a chain model of probability distribution (de Bonet et al., 1997) and it can be written as:

$$p(x) = \prod_{i=1}^{1} p(x_{\pi_1} \middle| x_{\pi_2}) p(x_{\pi_2} \middle| x_{\pi_3}) \ldots p(x_{\pi_{n-2}} \middle| x_{\pi_{n-1}}) p(x_{\pi_n})$$

where $\Pi=\{\pi_1,\pi_2,...,\pi_n\}$ is a permutation of the numbers $\{1,2,...,n\}$ used as an ordering for the pair wise conditional probabilities. At each iteration, the algorithm first tries to learn the linkage. Then, the algorithm uses a greedy algorithm to find a permutation Π that does not always give accurate model. Once the permutation Π is learnt, the algorithm estimates the pair wise conditional probabilities and samples them to get next set of solutions.

Combining Optimizers with Mutual Information Trees

The Combining Optimizers with Mutual Information Trees proposed by Baluja & Davies (1997, 1998) also uses pair-wise interaction among variables. The model of distribution used by this algorithm can be written as follows:

$$p(x) = \prod_{i=1}^{1} p(x_i \middle| x_j)$$

where, x_j is known as parent of x_i and x_i is known as a child of x_j. This model is more general than the chain model used by Mutual Information Maximization for input clustering as two or more variables can have a common parent.

Bivariate Marginal Distribution Algorithm

It was proposed by (Pelikan & Muhlenbein, 1999) as an extension to Univariate Marginal Distribution Algorithm. The model of distribution used by Bivariate Marginal Distribution Algorithm can be seen as an extension to the Combining Optimizers with Mutual Information Trees model and can be written as follows:

$$p(x) = \prod_{x_k \in Y} p(x_k) \prod_{x_i \in \{X|Y\}} p(x_i \middle| x_j)$$

where, $Y \subseteq X$ represents the set of root variables.

As a result, Bivariate Marginal Distribution Algorithm is a more generalised algorithm in this class and can cover both univariate interaction as well as bivariate interaction among variables.

Multivariate EDAs

The model of probability distribution becomes more complex than the one used by univariate and bivariate Estimation of Distribution Algorithms. Any algorithm considering interaction between variables of order more than two can be placed in this class. As a result, the complexity of constructing such model increases exponentially to the order of interaction making it infeasible to search through all possible models.

In what follows, we present the well-known works related to this category.

Extended Compact Genetic Algorithm

The Extended Compact Genetic Algorithm has been proposed by Harik (1999) as an extension to the Compact Genetic Algorithm. The model of distribution used in the Extended Compact Genetic Algorithm, is distinct from other previously described models as they only consider the marginal probabilities and do not include conditional probabilities. Also, it assumes that a variable appearing in a set of interacting variables cannot appear in another set. The model of distribution used by the Extended Compact Genetic Algorithm can be written as follows:

$$p(x) = \prod_{k \in m} p(x_k)$$

where, m is the set of disjoint subsets in n and $p(x_k)$ is the marginal probability of set of variables x_k in the subset k.

Factorised Distribution Algorithm

The Factorised Distribution Algorithm was proposed by Muhlenbein et al. (1999) as an extension to the Univariate Marginal Distribution Algorithm. The probability $p(x)$, for such linkage, can be expressed in terms of conditional probabilities between sets of interacting variables. In general, the Factorised Distribution Algorithm requires the linkage information in advance, which may not be available in a real world problem.

Bayesian Optimization algorithm

The Bayesian Optimization algorithm was proposed by Pelikan et al. (1999a). The probabilistic model $p(x)$ is expressed in terms of a set of conditional probabilities as follow:

$$p(x) = \prod_{i=1}^{n} p(x_i \,|\, \pi_i)$$

where, π_i is a set of variables having conditional interaction with x_i. Also no variable in π_i can have x_i or any children of x_i as their parent.

An extension to the Bayesian Optimization algorithm called hierarchical Bayesian Optimization algorithm has also been proposed by Pelikan & Goldberg (2000). The idea is to improve the efficiency of algorithm by using a Bayesian network with a local structure (Chickering et al., 1997) to model the distribution and a restricted tournament replacement strategy based on work of Harik (1994) to form the new population.

Estimation of Bayesian Network Algorithm

The Estimation of Bayesian Network Algorithm was proposed by Etxeberria & Larranaga (1999) and Larranaga et al., (2000) and also uses Bayesian networks as its model of probability distribution.

The algorithm has been applied for various optimisation problems, such as graph matching (Bengoetxea et al., 2000, 2001b,), partial abductive inference in Bayesian networks (de Campos et al., 2001), job scheduling problem (Lozano et al., 2001b), rule induction task (Sierra et al., 2001), travelling salesman problem (Robles et al., 2001), partitional clustering (Roure et al., 2001), Knapsack problems (Sagarna & Larranaga, 2001).

Learning Factorised Distribution Algorithm

The Learning Factorised Distribution Algorithm was proposed by Muhlenbein & Mahnig (1999b) as an extension to the Factorised Distribution Algorithm. The algorithm does not require linkage in advance. In each iteration, it computes a bayesian network and samples it to generate new solutions. The main steps in the Bayesian Optimization algorithm (BOA), the Estimation of Bayesian Network Algorithm (EBNA) and the Learning Factorised Distribution Algorithm (LFDA) procedures are shown in the following pseudo code:

BOA, EBNA and LFDA

1. Generate population *P* of *M* solutions
2. Select *N* promising solution from *P*.
3. Estimate a Bayesian network from selected solutions.
4. Sample Bayesian network to generate *M* new individual and replace *P*.
5. End the search if stop criteria are met, otherwise return to Step 2.

Markov Network Factorised Distribution Algorithm and Markov Network Estimation of Distribution Algorithm

The Markov Network Factorised Distribution Algorithm and the Markov Network Estimation of Distribution Algorithm were proposed by Santana (2003a, 2005). They used Markov network (Pearl, 1988; Li, 1995) as the model of distribution for $p(x)$. The first algorithm uses a technique called junction graph approach, while the second one uses a technique called Kikuchi approximation to estimate a Markov network.

PROBLEM STATEMENT

Manufacturing Cell Formation consists of grouping, or clustering, machines into cells and parts into families according to their similar processing requirements. The most known and efficient idea to achieve the objective of cell formation is to convert the initial Part Machine Incidence Matrix to a matrix that has a diagonal block structure. Among this process, entries with a '1' value are grouped to form mutually independent clusters, and those with a '0' value are arranged outside these clusters. Once a block diagonal matrix is obtained, machine cells and part families are clearly visible. However, the process engenders intercellular movements that require extra cost or time due to the presence of some parts that are processed by machines not belonging to its corresponding cluster. These parts are called Exceptional Elements. As a result, the objective of the block diagonalization is to change the original matrix into a matrix form minimizing Exceptional Elements and maximizing the goodness of clustering.

For cell formation problem, this matrix can be regarded as a binary matrix A which shows the relationship between any given m machines and p parts. Rows and columns represent respectively machines and parts. Each element in the matrix is usually represented by the binary entries a_{ij} where an entry of 1 indicates that a part i is processed by the corresponding machine j while an entry of 0 means a contrary situation. In Figure 1, we illustrate an (5×7) incidence matrix of King & Nakornchai (1982).

Figure 2 provides a block diagonal form for the initial matrix illustrated above. The obtained matrix has not any intercellular movement which means that it represents the optimal solution for the given matrix with 2 cells and 3 machines per cell.

In this chapter, we will deal with two efficient evaluation criteria namely the Grouping Efficacy (GE) and the Percentage of Exceptional Elements (PE). The Grouping Efficacy, proposed by Kumar & Chandrasekharan (1990), is considered one of

Figure 1. King & Nakornchai (1982) initial matrix

		Parts						
		1	2	3	4	5	6	7
Machines	1	0	1	0	1	1	1	0
	2	1	0	1	0	0	0	0
	3	1	0	1	0	0	0	1
	4	0	1	0	1	0	1	0
	5	1	0	0	0	0	0	1

Figure 2. A block diagonal matrix with no exceptional elements

		Parts						
		1	3	7	2	4	5	6
Machines	3	1	1	1	0	0	0	0
	2	1	1	0	0	0	0	0
	5	1	0	1	0	0	0	0
	4	0	0	0	1	1	0	1
	1	0	0	0	1	1	1	1

the best criteria which distinguish ill-structured matrices from well-structured ones when the matrix size increases and it is expressed as follows:

$$GE = \frac{e(X) - e_0(X)}{e(X) + e_v(X)}$$

Where:

$e_0(X)$: Number of Exceptional Elements in the solution X,
e: Number of 1's in the Part Machine Incidence Matrix,
$e_v(X)$: Number of voids in the solution X.

The second evaluation criterion is called the "Percentage of Exceptional Elements (*PE*)" is developed by Chan & Milner (1982) and expressed as follows:

$$PE = \frac{e_0(X)}{e} \times 100.$$

Some other performance measurements can be used to evaluate manufacturing cell design results. In what follows, we presents some of them.

The Grouping Efficiency which is developed by Chandrasekaran & Rajagopalan (1989). It expresses the goodness of the obtained solutions and depends on the utilization of machines within cells and inter-cell movements. This indicates that there are no voids and no exceptional elements in the diagonal blocks which imply a perfect clustering of parts and machines. Although grouping efficiency was widely used in the literature, it has an important limit which is the inability of discrimination of good quality grouping from bad one. Indeed, when the matrix size increases, the effect of 1's in the off-diagonal blocks becomes smaller, and in some cases, the effect of inter-cell moves is not reflected in grouping efficiency.

The Machine Utilization Index (MUI) which is defined as the percentage of the time that the machines within cells are being utilized most effectively and it is expressed as follows:

$$MUI = \frac{e}{\sum_i \left(m_i \times p_i\right)}$$

where m_i indicates the number of machines in cell i and p_i indicates the number of parts in cell i.

The Group technology efficiency which is defined as the ratio of difference between maximum number of inter-cell travels possible and number of inter-cell travels actually required by the system to the maximum number of inter-cell travels possible.

The Group efficiency which is defined as the ratio of difference between total number of maximum external cells that could be visited and total number of external cells actually visited by all parts to total number of maximum external cells that could be visited.

The Global efficiency is defined as the ratio of the total number of operations that are performed within the suggested cells to total number of operations in the systems.

PROPOSED EDA FOR MPCF PROBLEM (EDA-CF)

Solution Representation and Initial Population

Generally, for a Cell Formation Problem, a solution is represented by an m-dimensional vector $X=[x_1,x_2,\ldots,x_m]$ where x_i represents the corresponding assignment of the machine i to the specified cell. The problem consists in creating partitions of the set of the m machines assignments into a given number of cells. The created solutions must respect all the constraints defined in Section 3.3. We choose to generate the initial population randomly following a uniform distribution.

Selection

The goal is to allow individuals to be selected more often to reproduce. We adopt the truncated selection procedure to create new individuals: in each iteration, we select randomly P_1 individuals from the 50% of the best individuals in the current population. These P_1 individuals will be reproduced in the next generation using the probabilistic model to form new individuals.

Probabilistic Model and Creation of New Individuals

After the selection phase, a probabilistic model is applied to the P_1 selected individuals in order to generate new individuals.

The probabilistic model provides the assignment probability of the machine i to cell j and expressed in Box 1 where, $\varepsilon>0$ is a factor which guarantees that the model provides a probability $P_{ij}\neq 0$.

Replacement

The replacement represents the final step in our search procedure. It is based on the following idea: when a new individual is created, we compare it to the worst individual in the current population and we retain the best one.

Fitness Function

A fitness function is used for evaluating the aptitude of an individual to be kept or to be used for reproducing new individuals in the next generation. In the proposed algorithm, we used two fitness functions F_1 and F_2 to perform the objectives of minimizing the percentage of Exceptional Elements and maximizing the Grouping Efficacy respectively.

Let m_i be the number of machines assigned to the cell i. we define F_1 and F_2 as follows:

$F_1(X)=e_0(X)+Pen(X)$

Box 1.

$$P_{ij}=\frac{number\ of\ times\ where\ machine\ i\ appears\ in\ cell\ j\ +\ \varepsilon}{number\ of\ selected\ individuals\ +\ C\times\varepsilon}$$

and

$F_2(X)$=$GE(X)$-$Pen(X)$.

where:

$$Pen(X) = \alpha_1 \sum_{i=1}^{C} \max\left\{0, m_i - k_{\max}\right\} + \alpha_2 \sum_{i=1}^{C} \max\left\{0, 1 - m_i\right\}$$

expressed the distance between the solution X and the feasible space.

This penalty under-evaluate the fitness of solution X when X violate the constraint of the problem. i.e a penalty value is encountered either when the number of assigned machines exceeds the capacity of a cell or when machines are assigned to a number of cells that exceeds the fixed number of cells C.

Variable Neighborhood Search Algorithm

Variable Neighborhood Search is a recent meta-heuristic for combinatorial optimization developed by Mladenović & Hansen (1997). The basic idea is to explore different neighborhood structures and to change them within a local search algorithm to identify better local optima with shaking strategies. The main steps in this procedure are shown in the following pseudo code:

Variable Neighborhood Search

Select the set of neighborhood structures N_k,k={1,2,...,$n_{\max}$} that will be used in the search, find an initial solution X, choose a stopping condition.

Repeat the following steps until the stopping condition is met:

Set k=1

Repeat the following steps until all neighborhood structures are used:

1. Shaking: generate a point X' at random from k^{th} neighborhood of X $\left(X' \in N_k(X)\right)$
2. Local Search: apply some local search method with X' as initial solution; denote with X'' the obtained local optimum.
3. Move or not: if this local optimum X'' is better than the incumbent, or if some acceptance criterion is met, move there $(X \leftarrow X'')$, and set k=1 ; otherwise, set $k \leftarrow k+1$.

Local Search Procedure

Generally, obtaining a local minimum following a neighborhood structure does not imply that we obtain a local optimum following another one. For this reason, we choose to use two local search procedures which are based on two different neighborhood structures. The first neighborhood structure consists to select one machine and to insert it in a new cell. The second consists to select two machines from two different cells and to swap them.

Then, we apply these two local search procedures iteratively until there is no possible improvement to the current solution.

Shaking Phase

The main idea consists to define a set of neighbourhood structures that allow to obtain a distance equal to k between the solution X and the new neighbour solution X'. This distance can be defined by the number of differences between the two vectors X and X'. Then, we define N_k as the neighbourhood structure given by applying randomly k insertion moves.

COMPARATIVE STUDY

In order to show the competitiveness of the proposed EDA-CF algorithm, we provide in this section a comparative study with the well known approaches that treated Cell Formation problem. During all experiments, the proposed algorithm is coded using C++ and run on a computer Pentium IV with 3.2 GHz processor and 1GB memory.

Test Data Set

In order to evaluate the goodness of clusters obtained from the clustering heuristic for MPCF problem, 30 problems taken from the literature were tested. These data sets include a variety of sizes, a range from 5 machines and 7 parts to 40 machines and 100 parts, difficulties, and well structured and ill structured matrices.

For all instances, the initial matrix is solved by Estimation of Distribution Algorithm method and then improved by the Variable Neighborhood Search procedure. Then, the cells are formed and the machine layout in each cell is obtained optimally.

Table 1 shows the different problems and their characteristics. The columns illustrate respectively the sources of data sets, the problem size, the number of cells *C*, the maximum number per cell, k_{max} and the matrix density. All problems can be easily accessed from the references and they are transcribed directly from the original article they appeared. The appendix gives the block diagonal matrices for the improved solutions by the proposed algorithm. The maximum number of permissible cells *C* has been set equal to the best known number of cells as found in literature. The following equation expressed the density of the initial binary matrix and which informs about haw the one's elements are distributed inside the matrix.

$$\frac{\sum_{i}^{m}\sum_{j}^{n} a_{ij}}{m \times n}$$

Comparative Study

In this section, we evaluate the proposed algorithm by comparing it with the best results obtained by several well known algorithms respecting to the Grouping Efficacy and the Percentage of Exceptional Elements measures. In all tests, the proposed EDA-CF algorithm has proved its competitiveness against the best available solutions respecting to the same required number of cells.

As a stop condition to our algorithm, we fixed the maximal computational time to 5 seconds and the maximal number of iteration of Variable Neighborhood Search algorithm to 3.

The values of the following parameters are fixed as: ε=0,1; α_1=50; α_2=500; *P*=200 and P_1=3.

Comparison Respecting the Grouping Efficacy Measure

In this subsection we perform a comparative study with the best algorithm presented in the literature. These algorithms can be classified into two categories. The first category corresponds to the based population algorithm including Genetic Algorithm (GA) of Onwubolu & Mutingi (2001), Grouping Genetic Algorithm (GGA) of Brown & Sumichrast (2001), Evolutionary Algorithm (EA) of Gonçalves & Resende (2004) and Hybrid Grouping Genetic Algorithm (HGGA) of James et al. (2007). The second category represents the clustering based methods including ZODIAC of Chandrasekharan & Rajagopalan (1987), GRAFICS of Srinivasan & Narendran (1991), MST-Clustering Algorithm of Srinivasan (1994). Table 2 reports the results obtained by the proposed algorithm and these algorithms such that their results were taken from the original citations.

Table 1. Test problems from cellular manufacturing literature

No.	References	Size	C	k_{max}	Density
1	King & Nakornchai, 1982	5×7	2	4	0.400
2	Waghodekar & Sahu, 1984	5×7	2	5	0.5714
3	Seifoddini, 1989	5×18	2	12	0.5111
4	Kusiak & Cho, 1992	6×8	2	6	0.2987
5	Kusiak & Chow, 1987	7×11	3	4	0.2250
6	Boctor, 1991	7×11	3	4	0.2044
7	Seifoddini & Wolfe, 1986	8×12	3	5	0.6100
8	Chandrasekharan & Rajagopalan, 1986	8×20	3	9	0.2400
9	Chandrasekharan & Rajagopalan, 1986	8×20	2	11	0.3067
10	Mosier & Taube, 1985	10×10	3	4	0.3223
11	Chan & Milner, 1982	10×15	3	5	0.3646
12	Stanfel, 1985	14×14	5	6	0.1726
13	McCormick et al., 1972	16×24	6	7	0.2240
14	King, 1980	16×43	5	13	0.1831
15	Mosier & Taube, 1985	20×20	5	5	0.2775
16	Carrie, 1973	20×35	4	10	0.1957
17	Boe & Cheng, 1991	20×35	5	8	0.2186
18	Chandrasekharan & Rajagopalan, 1989 - 1	24×40	7	8	0.1365
19	Chandrasekharan & Rajagopalan, 1989 - 2	24×40	7	8	0.1354
20	Chandrasekharan & Rajagopalan, 1989 - 3	24×40	7	8	0.1437
21	Chandrasekharan & Rajagopalan, 1989 - 4	24×40	9	8	0.1365
22	Chandrasekharan & Rajagopalan, 1989 - 5	24×40	9	7	0.1375
23	Chandrasekharan & Rajagopalan, 1989 - 6	24×40	9	7	0.1365
24	McCormick et al., 1972	27×27	4	12	0.2977
25	Kumar & Vanelli, 1987	30×41	11	6	0.1041
26	Stanfel, 1985	30×50	12	7	0.1033
27	Stanfel, 1985	30×50	11	7	0.1113
28	King & Nakornchai,1982	36×90	9	27	0.0935
29	McCormick et al., 1972	37×53	2	35	0.4895
30	Chandrasekharan & Rajagopalan, 1987	40×100	10	6	0.1041

As seen in Table 2, in all the benchmark problems, the grouping efficacy of the solution obtained by the proposed method is either better than that of other methods or it is equal to the best one. We note that the solutions obtained by the GA method for problems 1, 7, 13, 24, 28 and 29 were not available. In five problems, namely 20, 21, 22, 23 and 24, the grouping efficacy of the solution obtained by the proposed method is better than that of all other methods. In other words, the proposed method outperforms all the other methods and the best solutions for these problems are reported in this paper for the first time. In eleven problems, namely 2, 3, 9, 15 and 17, the solution obtained by the proposed method is as good as the best solution available in the literature. In five problems, namely 4, 8, 10, 18 and 19, all the methods have obtained the same grouping efficacy.

Comparing with clustering methods, it is clear that the results obtained by the proposed algorithm are either equal or better than ZODIAC, GRAFICS and MST methods in all cases except for the problems 25 and 30. More specifically, the EDA-CF obtains for 6 (23%) problems values of the grouping efficacy that are equal to the best ones found in the literature by the three compared clustering methods and improves the values of the grouping efficacy for 19 (73%) problems.

Comparison Respecting the Percentage of Exceptional Elements Measure

Table 3 provides a comparison of the proposed algorithm against the best reached results available in literature. The comparison was done respecting to the Percentage of Exceptional Elements

Table 2. Summary of GE performance evaluation results

No	Size	C	GA	GGA	EA	HGGA	ZODIAC	GRAFICS	MST	EDA-CF	CPU
1	5×7	2	-	82.35	73.68	82.35	73.68	73.68	-	73.68	0.000
2	5×7	2	62.50	69.57	52.50	69.57	56.52	60.87	-	69.57	0.000
3	5×18	2	77.36	79.59	79.59	79.59	-	-	-	79.59	0.000
4	6×8	2	76.92	76.92	76.92	76.92	-	-	-	76.92	0.000
5	7×11	2	50.00	60.87	53.13	60.87	39.13	53.12	-	58.62	0.000
6	7×11	3	70.37	70.83	70.37	70.83	-	-	-	70.37	0.000
7	8×12	5	-	69.44	68.30	69.44	68.30	68.30	-	68.30	0.000
8	8×20	3	85.25	85.25	85.25	85.25	85.24	85.24	85.24	85.25	0.000
9	8×20	2	55.91	55.32	58.72	58.72	58.33	58.33	58.72	58.72	0.000
10	10×10	3	72.79	75.00	69.86	75.00	70.59	70.59	70.59	70.59	0.000
11	10×15	3	92.00	92.00	92.00	92.00	92.00	92.00	-	92.00	0.015
12	14×24	4	63.48	72.06	69.33	72.06	64.36	64.36	64.36	70.51	0.015
13	16×24	6	-	51.58	52.58	52.75	32.09	45.52	48.70	51.96	0.046
14	16×43	4	86.25	55.48	54.86	57.53	53.76	54.39	54.44	54.86	0.031
15	20×20	5	34.16	40.74	42.96	43.18	21.63	38.26	-	43.18	1.232
16	20×35	4	66.30	77.02	76.22	77.91	75.14	75.14	75.14	76.27	0.078
17	20×35	5	44.44	57.14	58.07	57.98	-	-	-	57.98	0.093
18	24×40	7	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	0.031
19	24×40	7	85.11	85.11	85.11	85.11	85.10	85.10	85.10	85.11	0.092
20	24×40	7	73.51	73.51	73.51	73.51	37.85	73.51	73.51	**76.97**	5.171
21	24×40	7	37.62	52.41	51.97	53.29	20.42	43.27	51.81	**72.92**	0.732
22	24×40	7	34.76	46.67	47.06	48.95	18.23	44.51	44.72	**53.74**	0.670
23	24×40	7	34.06	45.27	44.87	47.26	17.61	41.67	44.17	**48.95**	0.233
24	27×27	6	-	52.53	54.27	54.02	52.14	47.37	51.00	**54.98**	7.260
25	30×41	14	40.96	61.39	58.48	63.31	33.46	55.43	55.29	45.22	0.562
26	30×50	13	48.28	57.95	59.66	59.77	46.06	56.32	58.70	59.43	0.447
27	30×50	14	37.55	50.00	50.51	50.83	21.11	47.96	46.30	50.78	1.406
28	36×90	17	-	43.78	42.64	46.35	32.73	39.41	40.05	45.94	5.094
29	37×53	2	-	52.47	56.42	60.64	52.21	52.21	-	55.43	4.318
30	40×100	10	83.90	82.25	84.03	84.03	83.92	83.92	83.66	83.81	7.421

criteria. PE^a represents the best-known Percentage of Exceptional Elements found in the literature.

We note that the compared solutions for problems 15, 16, 17, 24, 26, 27, 28 and 29 were not available. The results shows that in all the benchmark problems, the number of exceptional elements of the solution obtained by the proposed method is either better than the best reached values or it is equal to the best ones. In 11 problems, namely 3, 6, 7, 12, 13, 14, 19, 20, 21, 22 and 23 the *PE* of the solution obtained by the EDA-CF is better than that of all other methods. In other words, the proposed method outperforms all the other methods. In nine problems, namely 1, 4, 5, 8, 9, 10, 11, 18 and 25, all the methods have obtained the same Percentage of Exceptional elements.

CONCLUSION

Cellular manufacturing is a production technique that leads to increase productivity and efficiency in the production floor. In this chapter, we have presented the first Estimation of Distribution Algorithm (EDA) method to solve the Machine Part Cell Formation Problem. Detailed numerical experiments have been carried out to investigate the EDAs' performance. Although the EDA approach does not require any problem-specific information, the use of sensible heuristics can improve the optimisation and speed up convergence. For this reason, we used the Variable Neighborhood Search (VNS) procedure in the improvement phase of the algorithm. The results from test cases presented here have shown that the proposed EDA-CF algorithm is very a competitive algo-

Table 3. Comparison between the obtained results and the best-known results respecting to the PE criterion

No.	size	C	Problem Source	PE	CPU	PE [a]
1	5×7	2	King & Nakornchai, 1982	0.000	0.000	0.000
2	5×7	2	Waghodekar & Sahu, 1984	0.150	0.000	0.125
3	5×18	2	Seifoddini, 1989	**0.000**	0.000	0.1957
4	6×8	2	Kusiak & Cho, 1992	0.0909	0.000	0.0909
5	7×11	2	Kusiak & Chow, 1987	0.1304	0.000	0.1304
6	711	3	Boctor, 1991	**0.0952**	0.000	0.1905
7	8×12	5	Seifoddini & Wolfe, 1986	**0.1714**	0.000	0.2857
8	8×20	3	Chandrasekharan & Rajagopalan, 1986	0.1475	0.000	0.1475
9	8×20	3	Chandrasekharan & Rajagopalan, 1986	0.2967	0.000	0.2967
10	10×10	3	Mosier & Taube, 1985	0.000	0.000	0.000
11	10×15	3	Chan & Milner, 1982	0.000	0.015	0.000
12	14×24	4	Stanfel, 1985	**0.0328**	0.015	0.1639
13	16×24	8	McCormick et al., 1972	**0.3721**	0.031	0.4302
14	16×43	4	King, 1980	**0.2063**	0.031	0.2222
15	20×20	6	Mosier & Taube, 1985	0.3693	0.078	-
16	20×35	5	Carrie, 1973	0.1985	0.031	-
17	20×35	5	Boe & Cheng, 1991	0.1764	0.062	-
18	24×40	7	Chandrasekharan & Rajagopalan, 1989 - 1	0.000	0.031	0.000
19	24×40	7	Chandrasekharan & Rajagopalan, 1989 - 2	**0.0308**	0.451	0.0769
20	24×40	7	Chandrasekharan & Rajagopalan, 1989 - 3	**0.1087**	5.171	0.1527
21	24×40	7	Chandrasekharan & Rajagopalan, 1989 - 4	**0.0992**	0.732	0.1527
22	24×40	7	Chandrasekharan & Rajagopalan, 1989 - 5	**0.2652**	0.670	0.3740
23	24×40	7	Chandrasekharan & Rajagopalan, 1989 - 6	**0.2824**	0.233	0.4214
24	27×27	6	McCormick et al., 1972	0.2350	0.203	-
25	30×41	14	Kumar & Vanelli, 1987	0.1094	0.219	0.1094
26	30×50	13	Stanfel, 1985	0.2754	3.109	-
27	30×50	14	Stanfel, 1985	0.1225	0.406	-
28	36×90	17	King & Nakornchai,1982	0.1254	0.969	-
29	37×53	3	McCormick et al., 1972	0.000	0.109	-
30	40×100	10	Chandrasekharan & Rajagopalan, 1987	0.0907	7.421	0.0857

rithm comparing with the previously published metaheuristics applied to the same problem. It has been shown that the EDAs provide efficient and accurate solutions for the test cases. The results are promising and encourage further studies on other versions of the Group Technology problems where we can introduce sequence data, machine utilization and routings.

REFERENCES

Andrés, C., & Lozano, S. (2006). A particle swarm optimization algorithm for part–machine grouping. *Robotics and Computer-integrated Manufacturing*, *22*, 468–474. doi:10.1016/j.rcim.2005.11.013

Askin, R. G., Creswell, J. B., Goldberg, J. B., & Vakharia, A. J. (1991). A Hamiltonian path approach to reordering the part-machine matrix for cellular manufacturing. *International Journal of Production Research*, *29*, 1081–1100. doi:10.1080/00207549108930121

Baluja, S. (1994). *Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning*. (Technical Report CMU-CS, 94-163). Computer Science Department, Carnegie Mellon University.

Baluja, S., & Davies, S. (1997). Using optimal dependency-trees for combinatorial optimization: Learning the structure of the search space. In *Proceedings of the 1997 International Conference on Machine Learning*.

Baluja, S., & Davies, S. (1998). *Fast probabilistic modeling for combinatorial optimization*. In AAAI-98.

Bengoetxea, E., Larranaga, P., Bloch, I., & Perchant, A. (2001b). Solving graph matching with EDAs using a permutation–based representation. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation*. Kluwer Academic Publishers. doi:10.1007/978-1-4615-1539-5_12

Bengoetxea, E., Larranaga, P., Bloch, I., Perchant, A., & Boeres, C. (2000). *Inexact graph matching using learning and simulation of Bayesian networks. An empirical comparison between different approaches with synthetic data*. In Workshop Notes of CaNew2000: Workshop on Bayesian and Causal Networks: From Inference to Data Mining, fourteenth European Conference on Artificial Intelligence, ECAI2000. Berlin.

Boctor, F. (1991). A linear formulation of the machine-part cell formation problem. *International Journal of Production Research*, *29*(2), 343–356. doi:10.1080/00207549108930075

Brown, E., & Sumichrast, R. (2001). CF-GGA: A grouping genetic algorithm for the cell formation problem. *International Journal of Production Research*, *36*, 3651–3669. doi:10.1080/00207540110068781

Burbidge, J. L. (1963). Production flow analysis. *Production Engineering*, *42*, 742–752. doi:10.1049/tpe.1963.0114

Caux, C., Bruniaux, R., & Pierreval, H. (2000). Cell formation with alternative process plans and machine capacity constraints: A new combined approach. *International Journal of Production Economics*, *64*(1-3), 279–284. doi:10.1016/S0925-5273(99)00065-1

Chan, H. M., & Milner, D. A. (1982). Direct clustering algorithm for group formation in cellular manufacture. *Journal of Manufacturing Systems*, *1*, 65–75. doi:10.1016/S0278-6125(82)80068-X

Chandrasekharan, M. P., & Rajagopalan, R. (1986a). MODROC: An extension of rank order clustering for group technology. *International Journal of Production Research*, *24*(5), 1221–1264. doi:10.1080/00207548608919798

Chandrasekharan, M. P., & Rajagopalan, R. (1987). ZODIAC: An algorithm for concurrent formation of part-families and machine-cells. *International Journal of Production Research*, *25*(6), 835–850. doi:10.1080/00207548708919880

Chandrasekharan, M. P., & Rajagopalan, R. (1989). Groupability: Analysis of the properties of binary data matrices for group technology. *International Journal of Production Research*, *27*(6), 1035–1052. doi:10.1080/00207548908942606

Chickering, D., Heckerman, D., & Meek, C. (1997). A Bayesian approach to learning Bayesian networks with local structure. In *Proceedings of Thirteenth Conference on Uncertainty in Artificial Intelligence,* (pp. 80–89). (Technical Report MSRTR- 97-07), Microsoft Research, August, 1997.

De Bonet, J., Isbell, C. L., & Viola, P. (1997). MIMIC: Finding optima by estimating probability densities. *Advances in Neural Information Processing Systems*, *9*, 424–430.

De Campos, L. M., Gamez, J. A., Larranaga, P., Moral, S., & Romero, T. (2001). Partial abductive inference in Bayesian networks: An empirical comparison between GAs and EDAs. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation*. Kluwer Academic Publishers. doi:10.1007/978-1-4615-1539-5_16

Etxeberria, R., & Larranaga, P. (1999). Optimization with Bayesian networks. In *Proceedings of the Second Symposium on Artificial Intelligenc*e. Adaptive Systems. CIMAF 99, (pp. 332-339). Cuba.

Gonçalves, J., & Resende, M. (2004). An evolutionary algorithm for manufacturing cell formation. *Computers & Industrial Engineering*, *47*, 247–273. doi:10.1016/j.cie.2004.07.003

Goncalves, J. F., & Resende, M. (2002). *A hybrid genetic algorithm for manufacturing cell formation. Technical report*. Rapport.

Gonzalez, C., Lozano, J. A., & Larranaga, P. (2002). Mathematical modelling of UMDAc algorithm with tournament selection: Behaviour on linear and quadratic functions. *International Journal of Approximate Reasoning*, *31*(3), 313–340. doi:10.1016/S0888-613X(02)00092-0

Harik, G. (1994). *Finding multiple solutions in problems of bounded difficulty*. Tech. Rep. IlliGAL Report No. 94002, University of Illinois at Urbana-Champaign, Urbana, IL.

Harik, G. (1999). *Linkage learning via probabilistic modeling in the ECGA*. Tech. Rep. IlliGAL Report No. 99010, University of Illinois at Urbana-Champaign.

Harik, G., Lobo, F., & Goldberg, D. E. (1998). *The compact genetic algorithm*, (pp. 523-528). (IlliGAL Report No. 97006).

James, T. L., Brown, E. C., & Keeling, K. B. (2007). A hybrid grouping genetic algorithm for the cell formation problem. *Computers & Operations Research*, *34*, 2059–2079. doi:10.1016/j.cor.2005.08.010

Joines, J. A., Culbreth, C. T., & King, R. E. (1996). Manufacturing cell design: An integer programming model employing genetic algorithms. *IIE Transactions*, *28*(1), 69–85. doi:10.1080/07408179608966253

Kaparthi, S., Suresh, N. C., & Cerveny, R. P. (1993). An improved neural network leader algorithm for part-machine grouping in group technology. *European Journal of Operational Research*, *69*, 342–355. doi:10.1016/0377-2217(93)90020-N

Khator, S. K., & Irani, S. A. (1987). Cell formation in group technology: A new approach. *Computers & Industrial Engineering*, *12*, 131–142. doi:10.1016/0360-8352(87)90006-4

King, J. R. (1980). Machine-component grouping formation in group technology. *International Journal of Management Science*, *8*(2), 193–199.

King, J. R., & Nakornchai, V. (1982). Machine-component group formation in group technology: Review and extension. *International Journal of Production Research*, *20*(2), 117–133. doi:10.1080/00207548208947754

Kumar, K. R., & Chandrasekharan, M. P. (1990). Grouping efficacy: A quantitative criterion for block diagonal forms of binary matrices in group technology. *International Journal of Production Research*, *28*(2), 233–243. doi:10.1080/00207549008942706

Kusiak, A. (1987). The generalized group technology concept. *International Journal of Production Research*, *25*, 561–569. doi:10.1080/00207548708919861

Kusiak, A., & Chow, W. S. (1987). Efficient solving of the group technology problem. *Journal of Manufacturing Systems*, *6*(2), 117–124. doi:10.1016/0278-6125(87)90035-5

Larranaga, P., Etxeberria, R., Lozano, J. A., & Pena, J. M. (2000). Combinatorial optimization by learning and simulation of Bayesian networks. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, (pp. 343–352). Stanford.

Li, S. Z. (1995). *Markov random field modeling in computer vision*. Springer-Verlag.

Lozano, J. A., Sagarna, R., & Larranaga, P. (2001b). Solving job scheduling with estimation of distribution algorithms. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation* (pp. 231–242). Kluwer Academis Publishers. doi:10.1007/978-1-4615-1539-5_11

Lozano, S., Adenso-Diaz, B., Eguia, I., & Onieva, L. (1999). A one step tabu search algorithm for manufacturing cell design. *The Journal of the Operational Research Society*, *50*, 509–516.

Mahdavi, I., Paydar, M. M., Solimanpur, M., & Heidarzade, A. (2009). Genetic algorithm approach for solving a cell formation problem in cellular manufacturing. *Expert Systems with Applications*, *36*, 6598–6604. doi:10.1016/j.eswa.2008.07.054

McAuley, J. (1972). Machine grouping for efficient production. *Production Engineering*, *51*(2), 53–57. doi:10.1049/tpe.1972.0006

McCormick, W. T. Jr, Schweitzer, P. J., & White, T. W. (1972). Problem decomposition and data reorganization by a cluster technique. *Operations Research*, *20*(5), 993–1009. doi:10.1287/opre.20.5.993

Mladenoviç, N., & Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, *24*, 1097–1100. doi:10.1016/S0305-0548(97)00031-2

Muhlenbein, H. (1998). The equation for response to selection and its use for prediction. *Evolutionary Computation*, *5*(3), 303–346. doi:10.1162/evco.1997.5.3.303

Muhlenbein, H., & Mahnig, T. (1999b). FDA - A scalable evolutionary algorithm for the optimization of additively decomposed functions. *Evolutionary Computation*, *7*, 353–376. doi:10.1162/evco.1999.7.4.353

Muhlenbein, H., Mahning, T., & Ochoa, A. (1999). Schemata, distributions and graphical models in evolutionary optimization. *Journal of Heuristics*, *5*, 215–247. doi:10.1023/A:1009689913453

Muhlenbein, H., & Paaß, G. (1996). From recombination of genes to the estimation of distribution. Binary parameters. *Lecture Notes in Computer Science, 1411*. *Parallel Problem Solving from Nature, PPSN*, *IV*, 178–187.

Onwubolu, G. C., & Mutingi, M. (2001). A genetic algorithm approach to cellular manufacturing systems. *Computers & Industrial Engineering*, *39*(1–2), 125–144. doi:10.1016/S0360-8352(00)00074-7

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Palo Alto, CA: Morgan Kaufman Publishers.

Pelikan, M., Goldberg, D. E., & Cantu-Paz, E. (1999a). BOA: The Bayesian optimization algorithm. In Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Pelikan, V., & Goldberg, D. E. (Eds.), *Hierarchical problem solving by the Bayesian optimization algorithm. IlliGAL Report No. 2000002*. Urbana, IL: Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign.

Pelikan, P., & Muhlenbein, H. (1999). The bivariate marginal distribution algorithm. In Roy, R., Furuhashi, T., & Chandhory, P. K. (Eds.), *Advances in soft computing-engineering design and manufacturing* (pp. 521–535). London, UK: Springer.

Robles, V., de Miguel, P., & Larranaga, P. (2001). Solving the travelling salesman problem with estimation of distribution algorithms. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation*. Kluwer Academic Publishers.

Roure, J., Sanguesa, R., & Larranaga, P. (2001). Partitional clustering by means of estimation of distribution algorithms. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation*. Kluwer Academic Publishers.

Sagarna, R., & Larranaga, P. (2001). Solving the knapsack problem with estimation of distribution algorithms. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation*. Kluwer Academis Publishers.

Santana, R. (2003a). A Markov network based factorized distribution algorithm for optimization. *Proceedings of the 14th European Conference on Machine Learning (ECMLPKDD 2003); Lecture Notes in Artificial Intelligence, 2837*, (pp. 337–348). Berlin, Germany: Springer-Verlag.

Santana, R. (2005). Estimation of distribution algorithms with Kikuchi approximation. *Evolutionary Computation*, *13*, 67–98. doi:10.1162/1063656053583496

Sierra, B., Jimenez, E., Inza, I., Larranaga, P., & Muruzabal, J. (2001). Rule induction using estimation of distribution algorithms. In Larranaga, P., & Lozano, J. A. (Eds.), *Estimation of distribution algorithms. A new tool for evolutionary computation*. Kluwer Academic Publishers. doi:10.1007/978-1-4615-1539-5_15

Sofianopoulou, S. (1997). Application of simulated annealing to a linear model for the formation of machine cells in group technology. *International Journal of Production Research*, *35*, 501–511. doi:10.1080/002075497195876

Solimanpur, M., Vrat, P., & Shankar, R. (2003). Ant colony optimization algorithm to the inter-cell layout problem in cellular manufacturing. *European Journal of Operational Research*, *157*(3), 592–606. doi:10.1016/S0377-2217(03)00248-0

Srinivasan, G. (1994). A clustering algorithm for machine cell formation in group technology using minimum spanning trees. *International Journal of Production Research*, *32*, 2149–2158. doi:10.1080/00207549408957064

Srinivasan, G., & Narendran, T. T. (1991). GRAFICS - A non hierarchical clustering-algorithm for group technology. *International Journal of Production Research*, *29*(3), 463–478. doi:10.1080/00207549108930083

Stawowy, A. (2006). Evolutionary strategy for manufacturing cell design. *OMEGA: The International Journal of Management Science*, *34*(1), 1–18. doi:10.1016/j.omega.2004.07.016

Venugopal, V., & Narendran, T. T. (1992a). A genetic algorithm approach to the machine component grouping problem with multiple objectives. *Computers & Industrial Engineering*, *22*(4), 469–480. doi:10.1016/0360-8352(92)90022-C

Xu, H., & Wang, H. P. (1989). Part family formation for GT applications based on fuzzy mathematics. *International Journal of Production Research*, *27*(9), 1637–1651. doi:10.1080/00207548908942644

Zhang, Q., Sun, J., Tsang, E., & Ford, J. (2004). Hybrid estimation of distribution algorithm for global optimisation. *Engineering Computations*, *2*(1), 91–107. doi:10.1108/02644400410511864

Zhao, C., & Wu, Z. (2000). A genetic algorithm for manufacturing cell formation with multiple routes and multiple objectives. *International Journal of Production Research*, *38*(2), 385–395. doi:10.1080/002075400189473

APPENDIX

Table 4. Problem 20

	4	16	7	14	23	24	9	10	17	2	5	11	19	6	8	12	15	18	3	20	1	13	21	22
8	1	1																						
19	1	1																						
21	1	1																						
28	1																							
37		1															1							
38	1	1																						
39	1	1																						
3			1	1	1	1																		
25			1	1	1																			
32		1	1	1	1	1																		
6							1	1	1															
7							1	1	1															
20								1	1							1								
29							1		1															
40							1	1	1															
10										1	1	1	1										1	
13										1		1	1											
14										1	1	1	1											
22										1	1	1	1											
35											1	1	1											
36											1	1	1											
4														1	1	1	1	1						
5											1			1	1	1	1	1						
18														1	1	1	1	1						
26								1						1	1		1	1						
27														1	1	1	1	1						
30														1	1	1		1						
2						1													1	1				

continued on following page

Table 4. Continued

	4	16	7	14	23	24	9	10	17	2	5	11	19	6	8	12	15	18	3	20	1	13	21	22
11																			1	1				
12																			1	1				
15																			1	1				
23																			1	1				
24																1			1	1				
31									1										1	1				
34																			1	1				
1																					1	1	1	1
9																					1	1	1	1
16																					1	1	1	1
17																					1	1		1
33										1											1	1	1	1

Table 5. Problem 21

	3	20	6	8	12	15	18	1	13	21	22	2	5	11	19	4	16	9	10	17	7	14	23	24
2	1	1																						1
11	1	1																						
12	1	1												1										
15	1	1																				1		
23	1	1																						
24	1	1			1																			
31		1																		1				
34	1	1																						
4			1	1	1	1	1																	
5			1	1	1	1	1						1											
18			1		1	1	1																	
26			1	1		1	1												1					

continued on following page

Table 5. Continued

	3	20	6	8	12	15	18	1	13	21	22	2	5	11	19	4	16	9	10	17	7	14	23	24
27			1	1	1	1	1																	
30			1	1	1		1				1													
1						1			1	1	1													
9								1	1	1	1													
16								1	1	1	1											1		
17								1	1		1													
33								1	1	1	1	1						1						
10										1		1	1	1	1									
13												1		1	1									
14												1	1		1									
22												1	1	1	1									
35													1	1	1									
36												1	1	1	1									
8																1	1							
19		1						1								1	1							
21																1	1							
28																1								
37						1											1							
38			1													1	1							
39																1	1							
6																		1	1	1				
7																		1	1	1				
20					1														1	1				
29																		1		1				
40															1			1	1	1	1			
3																					1	1	1	1
25																					1	1	1	
32																	1				1	1	1	1

Table 6. Problem 22

	3	20	2	5	11	19	6	8	12	15	18	1	13	21	22	10	4	16	7	24	9	17	14	23
2	1	1																		1				
11	1	1																						
12	1	1																						
15	1	1																						
23	1	1																						
24	1	1							1															
31		1																				1		
34	1	1																						
10			1	1	1	1								1										
13			1		1	1																		
14			1	1		1																	1	
22			1	1	1	1																		
35				1	1	1																		
36			1	1	1	1																		
4								1	1	1	1													
5				1			1	1	1	1	1													
18							1		1	1	1													
26							1	1		1	1					1								
27							1	1	1	1	1													
30							1	1	1		1				1									
1										1			1	1	1									
9												1	1	1										
16												1	1	1	1								1	
17												1	1		1									
33			1									1	1	1	1						1			
6																1					1			
20									1							1						1		
8																	1	1						

continued on following page

Table 6. Continued

	3	20	2	5	11	19	6	8	12	15	18	1	13	21	22	10	4	16	7	24	9	17	14	23
19		1										1					1	1						
21																	1	1						
28																	1							
37										1								1						
38							1										1	1						
39																	1	1						
32																		1	1	1				1
7																1					1	1		
29																					1	1		
40						1													1		1	1		
3																				1			1	1
25																			1				1	1

Table 7. Problem 23

	1	21	3	20	7	14	23	24	9	4	16	10	2	5	11	19	6	8	17	12	15	18	13	22
9	1	1																						
33	1	1							1				1											
2			1	1				1																
11				1						1												1		
12			1	1										1									1	
15			1	1											1									
23			1	1											1									
34			1	1														1						
3		1				1	1	1																
25				1	1	1	1																	
32	1				1		1	1			1													
6									1			1												

continued on following page

Table 7. Continued

	1	21	3	20	7	14	23	24	9	4	16	10	2	5	11	19	6	8	17	12	15	18	13	22
29									1										1					
39									1									1						
40					1				1							1								
8										1	1													
19	1			1						1	1													
21										1	1													
28										1														
37											1										1			
38										1							1							
20												1							1	1				
24			1									1								1				
10		1												1	1		1							
13							1						1		1	1								
14						1							1	1		1					1			
22													1	1	1					1				
35											1			1	1	1							1	
36											1		1	1		1								
26											1	1					1	1			1			
30																	1	1				1		1
7																			1					
31				1															1					
4																		1		1	1	1		
5														1				1		1	1	1		
18																	1			1		1		
27																1				1	1	1		
1		1																			1		1	1
16		1				1																	1	1
17	1																		1				1	1

Table 8. Problem 24

	6	8	18	2	15	19	9	4	7	14	13	22	1	21	23	24	17	3	20	10	12	16	5	11
4		1	1																		1			
5		1	1		1																		1	
18	1		1																		1			
26	1	1			1															1		1		
30	1	1	1									1												
38	1	1						1																
13				1		1			1						1									
14				1	1	1				1														
27			1		1	1															1			
36				1		1																	1	
6							1													1				
29							1										1							
39		1					1															1		
4						1	1		1															
11			1					1		1									1					
25									1	1									1					
28								1	1															
1					1						1	1		1										
16		1								1	1	1												
35						1					1												1	
9						1							1	1						1				
33				1			1						1	1										
3														1	1	1								
32									1						1	1						1		
7																	1							
17											1		1				1							
31													1				1		1					
2																1		1	1					

continued on following page

Table 8. Continued

	6	8	18	2	15	19	9	4	7	14	13	22	1	21	23	24	17	3	20	10	12	16	5	11
12							1				1				1			1	1				1	
15																		1	1					1
23																		1	1					1
34		1														1		1	1					
20																	1			1	1			
24																		1		1	1			
8								1														1		
19													1						1			1		
21								1														1		
37					1																	1		
10	1													1									1	1
22				1																	1		1	1

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 164-188, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 41
A LabVIEW-Based Remote Laboratory:
Architecture and Implementation

Yuqiu You
Morehead State University, USA

ABSTRACT

Current technology enables the remote access of equipment and instruments via the Internet. While more and more remote control solutions have been applied to industry via Local Area Networks (LANs), Wide Area Networks (WANs), and the Internet, there exist requirements for the applications of such technologies in the academic environment (Salzmann, Latchman, Gillet, and Crisalle, 2003). One typical application of remote control solutions is the development of a remote virtual laboratory. The development of a remote-laboratory facility will enable participation in laboratory experiences by distance students. The ability to offer remote students lab experiences is vital to effective learning in the areas of engineering and technology. This chapter introduces a LabVIEW-based remote wet process control laboratory developed for manufacturing automation courses. The system architecture, hardware integration, hardware and software interfacing, programming tools, lab development based on the system, and future enhancement are demonstrated and discussed in the chapter.

INTRODUCTION

As distance learning has progressed from basic television broadcasting into web-based Internet telecasting, it has become a very effective teaching tool (Kozono, Akiyama and Shimomura, 2002). Laboratory experiences are important for engineering and technology students to reinforce theories and concepts presented in class lectures. The development of a remote-laboratory facility will enable participation in laboratory experiences by distance students. The ability to offer remote students these lab experiences is vital to effective learning. The development of a remote

DOI: 10.4018/978-1-4666-1945-6.ch041

virtual laboratory is also motivated by the fact that presently, as never before, the demand for access to the laboratory facilities is growing rapidly in engineering and technology colleges. Being able to make the laboratory infrastructure accessible as virtual laboratories, available 24 hours a day and 7 days a week, goes far in addressing these challenges, and would also contribute to lowering the costs of operating laboratories. Additionally, remote virtual laboratories will provide the opportunity for students to explore the advanced technologies used in manufacturing remote control/monitor systems, and therefore, to prepare them for their future careers.

This chapter introduces a LabVIEW-based remote process control system, which is established to provide web based online virtual laboratory for an online computer-integrated manufacturing course. The physical setup of the system includes a wet process system, a FieldPoint control system with a NI cFP-2000 intelligent controller and eight I/O modules, a desktop computer, an Ethernet hub, and an Internet DLink camera. The wet process system is composed of three water tanks, pumps, discrete valves, continuous valves, temperature sensors, level sensors, and pressure transmitters. The software used for the system interfacing is LabVIEW 8.0 from National Instruments and the HyperText Markup Language (HTML). The desktop computer works as a control server as well as a web server for the system, providing a local interface for system control and maintenance and a remote interface for students to control and monitor the wet process through the Internet. The desktop computer, the intelligent controller, and the Internet camera communicate with each other through the Ethernet hub, and also connect to the Internet. All the sensors, valves, and pumps in the wet process system are wired to the I/O modules of the intelligent controller. Details of the system wiring will be examined later in this chapter. This system has been used in the lab of a computer-integrated manufacturing course for graduate students in the Manufacturing Technology option of the Technology Management program. Students are introduced to system integration, process control, LabVIEW FieldPoint programming, and the development of web-based manufacturing applications by involvement in the lab activities through the Internet.

This chapter explores the integration of the mechatronic equipment, computer software, and networking techniques to achieve a remotely controllable system. It demonstrates the development of a LabVIEW-based FieldPoint control system for a virtual laboratory. The implementation of the laboratory, future enhancement, and related researches are discussed in this chapter.

BACKGROUND

As in engineering and technology fields, the laboratory experiences associated with a technology curriculum are vital to understanding concepts (Saygin & Kahraman, 2004). They are also typically limited to a short group session each week due to time and space constraints. Increasingly practical and popular distance courses are hard-pressed to provide realistic lab experience at all. Simulation, which has seen increased use in education, is an especially valuable tool when it precedes instruction, but does not provide the problem-solving realism of actual hands-on experience (Deniz, Bulancak & Ozcan, 2003). Completing a project by remote operation of real equipment more nearly replicates problem solving as it would occur in the workplace and lends itself to teaching the processes and practice that are involved in true experimentation (Cooper, Donnelly & Ferreira, 2002).

With the rapid developments of computer networks and Internet technologies along with dramatic improvements in the processing power of personal computers, remote virtual laboratories are now a reality. In the early 1990's, the first remotely shared control system laboratory was proposed in the 1991 American Society of Engineering Education (ASEE) Frontiers in Education Conference. The system enabled sharing of laboratory data between universities using networked workstations. By its nature, automated manufacturing lends itself to remote access for education, and education that incorporates remote experimentation may better prepare students for the workplace of the future. With the development of standards for online lab sessions, the Accreditation Board for Engineering and Technology validated remote labs as educational tools (Carnevale, 2002). Saygin and Kahraman (2004) have successfully implemented lab exercises for manufacturing education and systems related courses using remote technology. Asumadu and Tanner (2006), who developed a remote wiring and measurement laboratory that utilizes a "virtual breadboard," acknowledge the flexibility and spontaneity of the tool which has potential for global access. Gurocak (2001) describes Programmable Logic Controller (PLC) and robot labs which are delivered via the remote lab concept by Internet connection to the PLCs and closed circuit TV of the labs. Thamma, Huang, Lou, and Diez (2004) integrated Computer-integrated Manufacturing equipment into a remote lab system via a Java based web site.

The virtual laboratory demonstrated in this chapter was developed by using LabVIEW FieldPoint control technology and a web-based application. The remote control panel provides real-time control and monitor function to remote clients and a live video of the real system operation. Remote clients can access the control panel from a web browser on their computers with the LabVIEW Runtime Engine installed as a plug-in. LabVIEW, developed by National Instruments, is a graphic programming language to build virtual instruments (VIs) for control systems. A VI developed in the LabVIEW environment provides an interface between a user and a control process. The main concept of such an interface is to provide a general view of the process and facilitate full control of the operations. LabVIEW is widely used in developing automatic control solutions in real world industries, research studies and academic laboratories. The locally controlled setup can be turned into a remotely controlled one by moving the user interface away from the physical setup with web-based functions. LabVIEW also provides advanced communication methods for the integration of LabVIEW VIs with other applications, such as ActiveX containers, File Input and Output, and.NET constructor nodes.

FieldPoint is a proprietary method for interfacing devices to computers developed by National Instruments, but it is very similar in principle to the concept of a fieldbus interfacing method used by many process control equipment suppliers. The idea of fieldbus grew out of the problem of interfacing hundreds or thousands of sensors and actuators to Programmable Logic Controllers (PLCs) and process control computers in large industrial plants. Instead of connecting each sensor or actuator to a central plant computer with hundreds or thousands of kilometers of wiring, the idea of fieldbus was to connect related groups of sensors and actuators to a local microcomputer that communicates with the central plant computer via an Ethernet local area network (LAN). The result was an enormous reduction in wiring and a corresponding increase in reliability. FieldPoint control in the LabVIEW environment is composed of four components, FieldPoint interface hardware, the FieldPoint Object Linking and Embedding for Process Control (OPC) server, the LabVIEW FieldPoint handler, and the Measurement and Automation Explorer (MAX). The FieldPoint interface hardware includes an intelligent controller, the I/O modules, and devices that connected to the modules. The FieldPoint OPC server is an invisible element of the software that can be invoked whenever the FieldPoint connection is setup with a LabVIEW application programmed for the control system. The communication with the FieldPoint unit does not necessarily occur at the precise instant when the application instructs it to happen. Instead, communication occurs both as a direct result of requests from the application and also as a result of the configuration for the intelligent controller. Any LabVIEW application that uses FieldPoint can be conveniently structured by incorporating all FieldPoint operations into a single module -- a FieldPoint handler. This module performs four different operations: initialization, close, read, and write. This handler directs all the communications between the FieldPoint unit and the LabVIEW application. The MAX provides an interface for the setup and configuration of the FieldPoint hardware. From the MAX interface, the LabVIEW application can locate and recognize the FieldPoint hardware, and the devices in the FieldPoint unit can be tested through data communication directly from the MAX.

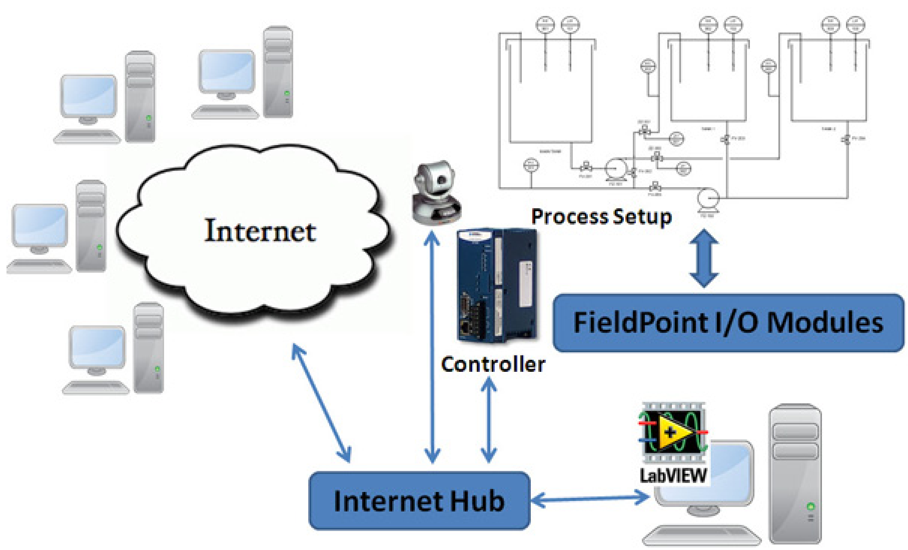
In the system introduced in this chapter, a virtual interface programmed in the LabVIEW graphical language provides a control panel for users to interact with the control process through FieldPoint Ethernet communication and the communication between the FP controller and I/O Modules. The remote laboratory introduced in this chapter provides an approach to implementing LabVIEW interfacing technology in manufacturing process control systems to provide remote virtual lab activities for students in a manufacturing engineering program. The development of this laboratory also has students explore the integration of computer and networking technologies into manufacturing control systems for higher flexibility and productivity. Researches and experiments related to web-based manufacturing control systems and remote virtual laboratories are being conducted based on this system.

SYSTEM OVERVIEW

Physical Setup of the System

The technology used for the development of the virtual laboratory system in the LabVIEW environment is FieldPoint. As mentioned earlier in this chapter, the physical setup of the system includes a wet process system, a FieldPoint control system with a NI cFP-2000 intelligent controller and eight I/O modules, a desktop computer, an Ethernet hub, and an Internet DLink camera. As shown in Figure 1, the intelligent controller, the Internet camera, and the desktop computer are connected to an

Figure 1. The system setup



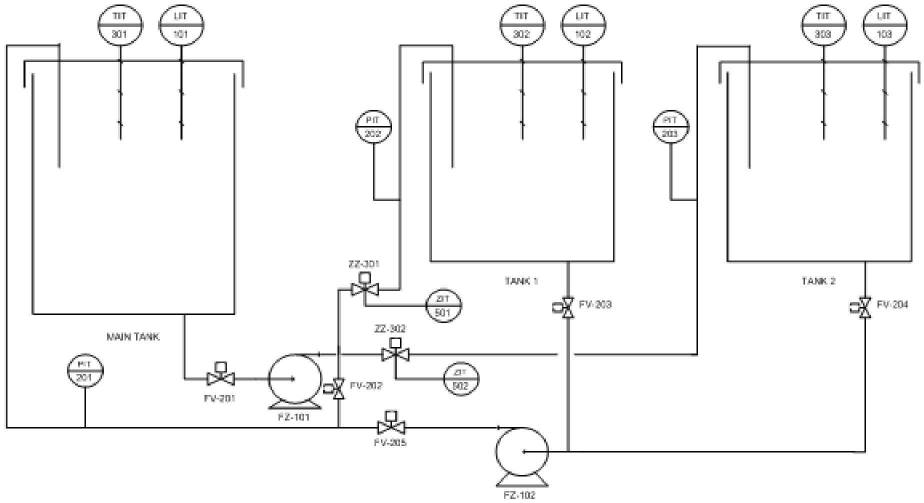
Internet hub, and then connected to the Internet. Clients (students) can access the remote control panel of the system over the Internet connection.

The wet process control system is comprised of three tanks, two pumps, five discrete valves, and two continuous valves, as shown in Figures 2 & 3. The sensors used in the system include three temperature sensors, three level sensors, one flow rate sensor, and two pressure sensors. Temperature sensors and level sensors were used to monitor each tank's water level and temperature. Two pressure sensors were installed to monitor the incoming flow pressure of tank 1 and tank 2. A flow sensor was installed to measure the incoming flow rate of the main tank. Figure 2 shows a picture of the real wet process system setup. In order to better demonstrate the level changes of the water in each tank, the water was dyed with green color. Also, a physical control panel with pushbuttons and a panel with light indicators was added to the system for local maintenance and control. The physical intelligent controller and its I/O modules can be seen on the picture as a blue panel mounted on the wall. Figure 3 demonstrates the components of the physical setup to give you a clearer idea of the system setup with pipeline connections running between tanks and the locations of all the system components.

All the values, pumps and sensors were wired to the Input/Output modules of the intelligent controller. There are four Input/Output modules used in this system, which include the analog input module cFP-AI-110, the analog output module cFP-AO-200, the digital output module cFP-DO-400, and the temperature module cFP-RTD-124. The National Instruments cFP-AI-110 is an 8-channel single-ended input module for direct measurement of milli-volt, low voltage, or milli-ampere current signals from a variety of sensors and transmitters. It delivers filtered low-noise analog inputs with 16-bit resolution, and features overranging, HotPnP (plug-and-play) operation, and onboard diagnostics. The National Instruments cFP-AO-200 is an 8-channel analog output module for 4 to 20 mA and 0 to 20 mA current loops. The module includes open-circuit detection for wiring and sensor trouble-shooting and short- circuit protection for wiring errors. It features HotPnP operation so it is automatically detected and identified by the configuration software. The National Instruments cFP-DO-400 module features eight sourcing digital

Figure 2. Picture of the wet process control system

Figure 3. Physical setup diagram of the wet process system



output channels. Each channel is compatible with voltages from 5 to 30 VDC and can source up to 2 A per channel with a maximum of 9 A squared per module (The sum of the squares of the output currents from all eight channels must be no greater than 9). Each channel has an LED to indicate the channel on/off state. The module features 2300 V transient isolation between the output channels and the backplane. It also features HotPnP operation and is automatically detected and identified by the configuration software. The National Instruments cFP-RTD-124 is an 8-channel input module for direct measurement of 2 and 4-wire RTD temperature sensor signals. With current excitation, signal conditioning, double-insulated isolation, input noise filtering, and a high-accuracy delta-sigma 16-bit analog-to-digital converter, it delivers reliable, accurate temperature or resistance measurements.

Table 1 provides a detailed list of the input and output devices in the physical setup of the system and the type of FieldPoint I/O module they were wired with. As shown in Table 1, the discrete valves and the pumps were wired to the digital output module; the continuous control valves were wired to the analog output module; the temperature sensors were wired to the RTD module for temperature readings; and the level sensors, the flow rate sensor, and the pressures sensors were wired to the analog input module.

LabVIEW Interfacing

As mentioned in the previous section, the devices on the wet process system including all the valves, pumps, and sensors, were wired to the four different Input/Output modules of the FieldPoint controller unit. All these devices were setup and configured through the Measurement & Automation software (MAX), and each of them was assigned a unique ID starting with an IP address, as shown in Table 1. The FieldPoint controller unit is recognized by the computer as a network node with its IP address. For the system demonstrated here, 139.102.29.56 was assigned to the controller. The four Input/Output modules then were recognized as different communication ports under this IP address. In this system, they are recognized as port 1, 2, 5, 7 respectively. Then, each device wired to the same Input/Output module was identified by a unique channel number. The unique ID for each device therefore is a combination of the IP

Table 1. I/O addressing of the control system

LabVIEW Addressing in Programming		Device #	Device Description
FP@139_102_29_56\cFP-DO-400@7 (Digital Output Module)	\Channel 0	FZ-101	Pump 1
	\Channel 1	FZ-102	Pump 2
	\Channel 2	FV-201	Discrete valve 1
	\Channel 3	FV-202	Discrete valve 2
	\Channel 4	FV-203	Discrete valve 3
	\Channel 5	FV-204	Discrete valve 4
	\Channel 6	FV-205	Discrete valve 5
FP@139_102_29_56\cFP-AO-200@5 (Analog Output Module)	\Channel 0	ZZ-301	Continuous control valve 1
	\Channel 1	ZZ-302	Continuous control valve 2
FP@139_102_29_56\cFP-RTD-124@2 (Temperature Module)	\Channel 0	TIT-301	Temperature Sensor 1
	\Channel 1	TIT-302	Temperature Sensor 2
	\Channel 2	TIT-303	Temperature Sensor 3
FP@139_102_29_56\cFP-AI-110@1 (Analog Input Module)	\Channel 2	LIT-101	Level Sensor 1
	\Channel 3	LIT-102	Level Sensor 2
	\Channel 4	LIT-103	Level Sensor 3
	\Channel 5	PIT-201	Flow Rate Sensor
	\Channel 6	PIT-202	Pressure Sensor 1
	\Channel 1	PIT-203	Pressure Sensor 2

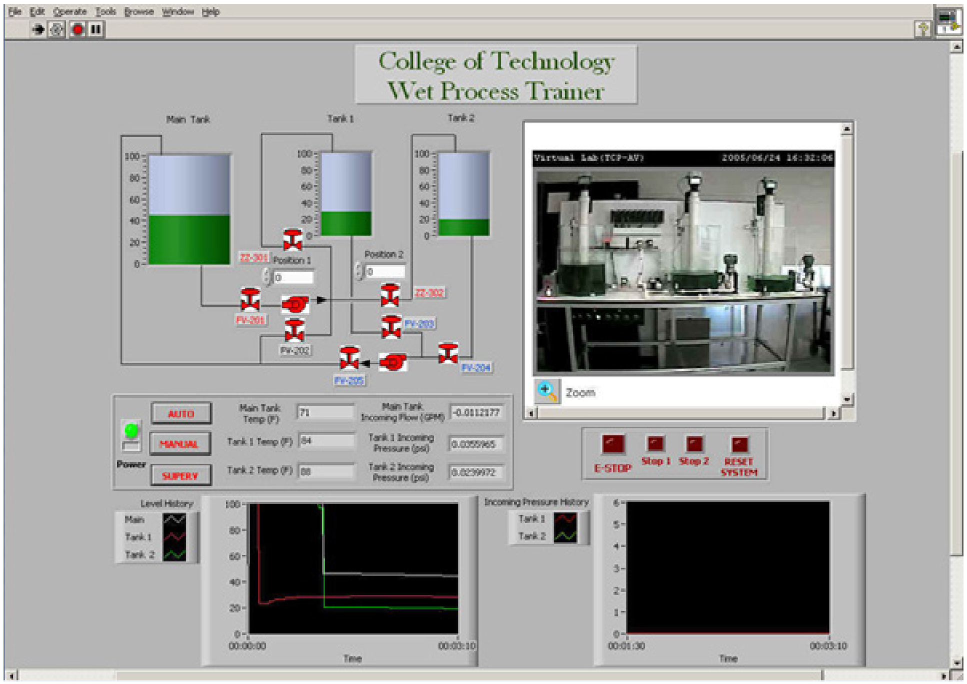
address (identifying the controller unit), the port number (identifying the specific module), and the channel number (identifying the device).

Once the physical system was connected and configured for the LabVIEW communication, a virtual interface programmed by using the LabVIEW graphical language can provide a control panel for users to interact with the control process on the local server as well as over the Internet from a client computer as shown in Figure 4. The virtual interface is called a virtual instrument (VI) in the LabVIEW environment.

As shown in Figure 4, this virtual interface has five major areas for users to interact with the real control process including a process simulator, a mode control panel with digital indicators, a stop button panel, a live video window, and waveform graphics for data tracking. The process simulator simulates the real wet control process by using control icons and indicator icons. In manual mode, users can control each individual device of the real process by clicking on the control icons, such as the valves and motors. Indicator icons will display the status of those devices by changing the color of icons to green or red. The mode control panel is used to change control mode, and can also provide real data readings from sensors. The stop button panel provides different buttons to stop major devices and the system itself. The current value of tank levels, temperatures, incoming flow rate for main tank, and incoming pressure for tank 1 and tank 2 are displayed by the graphic and digital indicators on the interface. A video window was integrated into the interface for users to monitor the real process through an Internet camera. Two waveform graphics windows provide history data tracking the temperature and incoming pressure of each tank.

Behind the Frontpanel of the virtual interface is the block diagram, programmed to provide the data flow, mathematic operations, and logic operations of the virtual instrument. The application

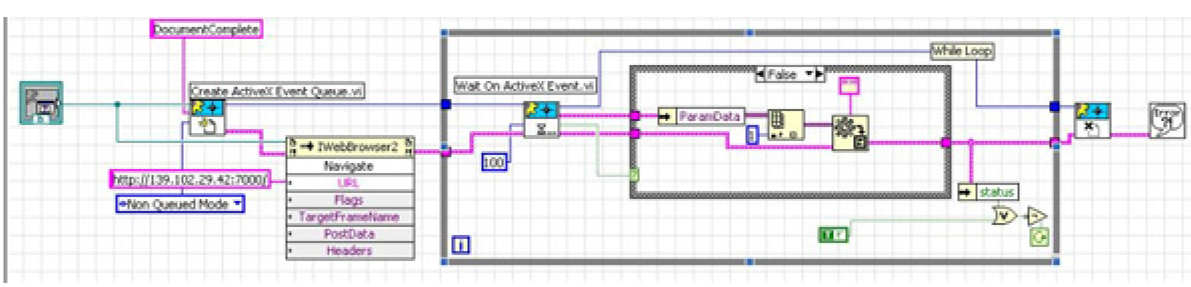
Figure 4. The virtual interface programmed in LabVIEW



programming utilized Case Structure functions to provide three different modes, auto mode, manual mode, and the supervision mode. A While Loop function was used to establish continuous data retrieving and command sending cycles from and to the physical wet process system. Part of the block diagram programmed for this interface is shown in Figure 5 which is a live video retrieving diagram from the Internet camera.

Real time control from the virtual interface on the real system is through the FieldPoint Ethernet communication and the communication between the FP controller and I/O Modules. FP controller, FB OPC Server, and FP manager are installed and configured through Measurement & Automation software (MAX). The communication between the FP controller and I/O Modules is similar among different types of network modules. The FP controller communicates with each module through the Ethernet module using the TCP/IP protocol. It uses the.iak file to determine which resource to communicate with. Each I/O module cycles through its internal routine of sampling all channels, digitizing the values and updating the values

Figure 5. Block diagram programmed in LabVIEW



on the module channel registers (buffer). This cycle time is set for each module and is specified as the all channel update rate. FieldPoint Ethernet communication uses an asynchronous communication architecture called event-driven communication. The network module automatically sends updates to a client when data changes. The server then caches the data from I/O modules and uses it to respond to read requests from the virtual interface. The network module scans all I/O channels with subscriptions to determine if a value has changed by comparing the current value to the cached value for each channel. If a change has occurred, the network module puts the difference between the two values in the transmit queue. The FP Server receives this information and sends an acknowledgement to the network module. The network module periodically sends and receives a time-synchronization signal so that it can adjust its clock and provide proper timestamping. When signals do not change over long periods of time, the client sends periodic re-subscribe messages to verify that the system is still online.

LabVIEW's architecture allows for easy integration of the laboratory environment for remote manipulation. The main concept of turning the locally controlled setup into a remotely controlled one is moving the user interface away from the physical setup. The local computer works as the web server as well as the control server. A number of clients can log onto the server, but only one user can be granted the control right. Other users can monitor the control process. They can monitor the process from their remote front panel (VI) while the one that has the control right can actually control the process from the panel. There is a waiting queue for users. When the control right is available, it will be granted to the next user in the queue. The remote client can be any computer with Internet access. The only tool that the client needs to use is a web browser with the LabVIEW Runtime Engine installed. The LabVIEW Runtime Engine is plug-in software provided by National Instruments to support the web application. Normally it is installed automatically on the client's machine the first time the user tries to view a front panel. The client can browse to the webpage integrated with the remote control panel by entering the Uniform Resource Locator (URL) address of the web server in the browser. The client only updates the screen and gets information from front panel interactions. The client cannot make changes to VIs. Execution happens only on the server machine.

The local server hosts a LabVIEW web server, which publishes the VI to the Internet. Through the LabVIEW Real Time Engine (RTE), the local server can communicate with the remote client. It controls the process according to the data from the remote control panel and sends the updated data back to the remote control panel. Remote clients are not required to have the whole LabVIEW software installed to view VIs for control and monitoring. They just require the LabVIEW RTE plug-in.

The security of the control system is ensured by management from the server side. On the server side, the user's permission to access the LabVIEW control panel is managed through editing the allowed list of IP addresses for clients. Also, access to the LabVIEW control panel can be limited to a specific domain or a group of domains. The virtual interface running on the server can be configured to be available to or be hidden from certain users. While in the process of remote control and monitoring, the IP address of the active client will be shown on the server. The lab instructor can always monitor the usage of the remote control panel, and make sure only authorized clients have access permissions. In the process of remote control, the lab instructor can take over the control right on the server side at any time in case of system malfunctions, user errors, or any unusual situations.

System Operation

Figure 6 shows the process simulator on the LabVIEW virtual interface. This process diagram simulates the real process of the system with controls/indicators corresponding to the components of the real system. Three tank indicators represent the main tank, tank 1 and tank 2 in the wet process trainer respectively. The green bar of each tank indicates the current water level which can vary from 0 percent to 100 percent. In this diagram, valves and pumps are controls for their corresponding parts in the real process. FV-201, FV-202, FV-203, FV-204, and FV-205 are controls for discrete valves. These five discrete valves and two pumps are controlled by ON/OFF Boolean signals; their status can be changed by a mouse-click on them. The color of each control represents the status of the control. Red represents OFF status while green represents ON status. ZZ-301 and ZZ-302 represent the two continuous control valves in the real process. The status of them is controlled by the digital control below the valve icons. The value of each digital control can be changed by clicking the arrows beside it from 0 to 100 with the increment of 10. The value of 0 represents closed status of the valve, and the value of 100 represents the totally open status of the valve. The color of ZZ-301 and ZZ-302 will be changed to green color when the value of their digital control is equal or greater than 10 to indicate an ON status. Otherwise, it will be changed to red indicating an OFF status. This process simulator provides a direct visual view of the whole wet process system for students to understand the process and identify the function of each component of the system. It can be used in the manual control mode for control testing and in the supervision mode for system maintenance and trouble-shooting.

Figure 7 displays the part of the virtual interface with power control, mode controls and digital indicators. The power control is used to turn on/off the system, which will be in green color when the system is running. The six indicators to the right of the mode buttons are digital

Figure 6. Process simulator

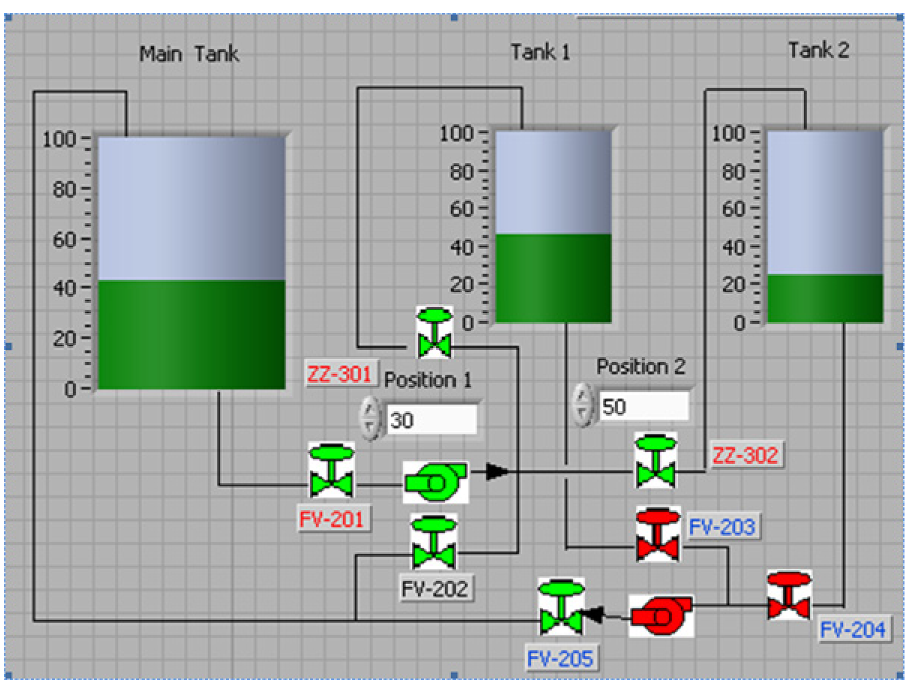


Figure 7. Mode controls and digital indicators

indicators displaying current values of tank levels, liquid temperatures, incoming flow rate, and incoming pressures.

The buttons with red labels are used to select a control mode for the system operation. This virtual interface provides three different modes for process control which include supervision mode, manual mode, and auto mode. Clients (students) are assigned different control capabilities when different modes are selected.

In supervision mode, all the valves and pumps in the wet process system can be turned on and off by users disregarding the readings from the sensors. This mode can be enabled only for maintenance and trouble-shooting purposes, and it is not available to remote clients for security reasons. In manual mode, the status change of valves and pumps will depend on both the current situation of the system and the commands from the user. For example, if the water level of the main tank is lower than 20 percent or valve FV-201 is closed, pump 1 cannot be activated even if the user intends to do so by clicking on the pump icon on the process simulator of the virtual interface. When certain conditions are met from the sensor readings (which means it is a safe situation), the user can manually control any device of the wet process system. This mode is available for both local and remote users. It helps users to test device status, get familiar with the control interface, and adjust control parameters when necessary. In auto mode, the system demonstrates a fully automated control process to users, depending on the water levels of each tank. The user can only control continuous control valves by changing the percentage values without changing the on/off status of the valves. The user can adjust the percentage of the continuous valves by clicking the toggle switch beside the digital control or typing values in the digital controls directly. These three modes provide flexibility for students to explore the wet control process, and also ensure the security to protect the system physical setup.

The LabVIEW interface panel provides three emergency stop buttons and one reset button in the stop button panel. E-Stop button will stop the whole system when pressed. Stop 1 and Stop 2 buttons will disable pump 1 and pump 2 respectively when pressed. The Reset button is used only in auto mode to reset the system when the main tank level reaches its limits. This stop button panel is available for both local and remote users. Lab instructor can also disable the system by clicking the stop button on the LabVIEW window from the server side, or press the emergency stop button located on the physical system in any emergency situations.

The live video window integrated in the LabVIEW virtual interface displays the live video from the Internet camera to remote users. Remote users can view the liquid levels and the light indicators clearly from a remote interface. There are seven green light indicators on the indicator panel mounted on the physical system. The seven green lights were wired to the five discrete valves and two pumps to indicate the On/Off status. This will help remote users to compare the status of devices on the virtual interface and the ones on the physical system when necessary for their operation. This is a great tool to help remote users observe the operations of the real system.

LABORATORY IMPLEMENTATION

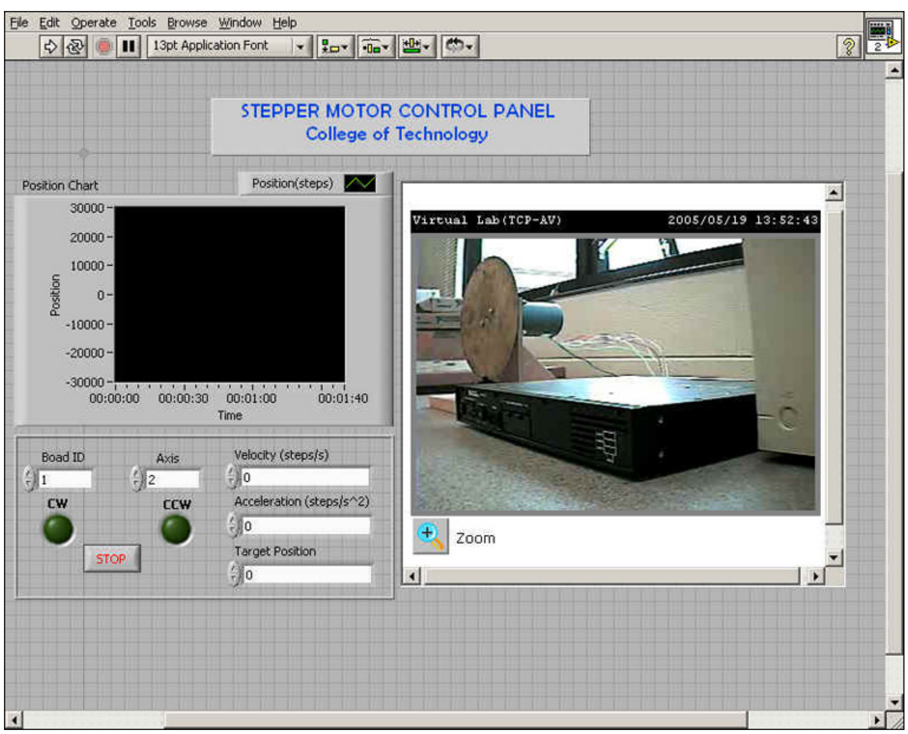
There are six labs that have been developed and implemented in the computer-integrated manufacturing course as part of the required lab activities.

Lab 1: Introduction to the LabVIEW Environment

The purpose of this lab is to introduce the LabVIEW graphic programming environment to students. A simple single-axis motor control system is used to help students get familiar with the LabVIEW front panel and block diagram. First, students will remotely access the single-axis motor control system developed in LabVIEW to control and monitor a stepper motor over the Internet. The LabVIEW interface for this motor control system is shown below in Figure 8.

A LabVIEW program consists of two parts, the front panel and the block diagram. The front panel is used to design graphic interfaces; a control palette provides various controls and indicators to be used for a control interface. The program which runs behind the graphic interface is called the block diagram. A function palette associated with this block diagram provides all kinds of functions and operations. In the second part of this lab, students are required to install a LabVIEW student version on their computers. Students will open the LabVIEW front panel window and block diagram window to explore each control and indicator used in this simple program, so that they can get familiar with the LabVIEW programming environment.

Figure 8. Single-axis stepper motor control interface



Lab 2: Programming the Motion Control

The purpose of this lab is to help students understand the major components in a motion control system and functions of basic motion control VIs, and gain skills in programming motion control systems. Students are required to develop a simple single-axis motor control program by following instructions provided by the instructor, send the program to the lab instructor, and run their own program for remote motor control. The board ID and the axis number will be assigned to each student for their programming process. The lab instructor configures the motor and monitors students' remote operations on the control server.

Lab 3: Introduction to Remote Process Control Using FieldPoint

The purpose of this lab is to help students understand the integration of input and output devices in FieldPoint, examine the available devices and technologies used for remote process control applications, and explore the implementation of remote process control applications. Students are required to operate the virtual process control system through the Internet (as shown in Figure 2), examine each mode available on the virtual interface, and understand the mechanism for the system integration. The lab instructor will monitor the control process while students are using the virtual control interface to access the wet process setup in the lab.

Lab 4: Programming a Simple Process Control Program in FieldPoint

The purpose of this lab is to gain skills in design and to program a process control system with digital input/output signals and Boolean operations. Students are required to design a process control system with part of the components available in the physical system setup including valves and pumps, send their program to the lab instructor, and test their program through the Internet after the program is loaded into the control server.

Lab 5: Programming for Remote Measurements

The purpose of this lab is to help students understand the mechanism for retrieving analog data remotely, examine the available devices and technologies for remote data acquisitions, and gain skills in developing a remote data retrieving system. Students are required to design a LabVIEW VI for remote analog data retrieving from temperature sensors, level sensors, and pressure sensors. These sensors are already installed as part of the wet process control system, as shown in Figure 3 and Table 1. The lab instructor will send the data sheets of the sensors to students. Students will then design the interface, and program the block diagram by using the technical data provided by the lab instructor to retrieve signals from sensors and display correct data on their program interfaces. Students will send their programs to the lab instructor and test their programs after the program is loaded into the system.

Lab 6: Design a Virtual Manufacturing Work Cell

This lab is to encourage students to apply their knowledge, skills and experiences gained from the lectures and lab activities to design a virtual manufacturing cell for remote control and monitoring. Students are required to use sensors for measurements and use motors to simulate machine status. The machines in the manufacturing cell include three Computer Numeric Controlled (CNC) machines, three industrial robots, and one conveyor system. Based on what they have learnt, students need to integrate machines and devices into a connected network by using a FieldPoint system, assign an Input/Output address to each device and machine, design the control interface, and program the block diagram. Then students will send their programs to the lab instructor, test their programs by operator and monitor the sensors, valves, and motors of the process control system through the Internet.

The educational value of these online lab activities has been assessed through students' feedback. It is shown that most students find these labs are very interesting, convenient to access and easy to follow. They consider the lab necessary for them to understand the concepts of mechatronic system integration, remote control, and Human-Machine Interface (HMI). Students gain experience by exploring, operating, and programming the system. In addition to helping students understand the concepts and principles of remote control applications in manufacturing, these lab activities provide the following major benefits collected from students' feedback.

- Hands-on experience with LabVIEW programming.
- Great hands-on experience with online control and monitoring.
- Broader view on the future of industrial networking in implementing computer-integrated manufacturing.
- A convenient way to access lab facilities.
- Flexible schedule to work on lab activities.

However, some of the students did mention that the time delays in the control process have caused problems for their remote operation, and the programming difficulties with remote measurements have made them frustrated when testing is not available until the program has been completed. These problems could be solved in the future with system enhancement by implementing more web-based applications. Also, the online laboratory is not available 24/7 for online students now due to the safety and security concerns. But it does provide convenient access to labs for online students. They can schedule their lab activities in evenings and weekends with the lab instructor when either the lab instructor or a lab assistant can sit by the server or monitor the process through the Internet.

ISSUES OF RELEVANCE TO THE LABORATORY

There are several issues related to this virtual laboratory that will affect the future development of the laboratory according to feedback from students, lab instructors, and faculty. The issues include the influence of network bandwidth on information transmission for remote control, the user management system, and the limited functions of the LabVIEW web server to support online programming.

In this virtual process control system, time delays exist in data transmission especially when the client accesses through a dial-in network connection. Obviously it is caused by the different bandwidth of networks. In the development of the remote laboratory, not only parameter and administrative data but also audio and video data need to be transmitted via network connections. Web cameras will bring live images of the physical setups to remote clients. Undoubtedly, it is critical to use the available bandwidth efficiently in data transmission in the project; otherwise, time delays will mar the whole remote experimentation execution. Several networking techniques can be used in solving this problem. For example, setting respective data priorities for the transmission rate of different data can ensure critical data is transmitted without delay. Another technique that can be used is data compression, but it involves a trade-off because of the additional delay resulted from the compression and decompression processes. This delay should be kept much smaller than the transmission delay. Data compression is especially useful for audio/video transmission which involves a huge amount of data. At the same time, the server needs to adapt to different bandwidth requirements of remote clients. For example, some might be on the same local LAN with the server on the same campus while others may be connected from home using a dial-up line.

When implemented in the lab of the graduate course, there were only 12 students in the class. There was not any complaint from students about the access method. If implemented in classes with more students, user management becomes an issue. Not all students would like to wait in the queue for their online lab activities. A user management system will be developed using Visual Studio 2005 to integrate an interface and an ACCESS 2007 database, and communicate with the LabVIEW web server to realize a user reservation system. This will increase the flexibility for students allowing them to login and schedule their lab activity online. It will also provide a more secure way to manage users with permission assignments.

The function of the LabVIEW web server is the second issue in the remote virtual laboratory. In this remote control process, the LabVIEW web server publishes the VIs to the Internet, but clients can only update and get information from front panel interactions. Clients cannot make changes to VIs directly from the remote interface. In order to let students to learn and practice programming in the LabVIEW environment, clients must have the capability to re-program the process and re-download their programs to the controller for testing purpose. This takes time for students to complete one program-test cycle, and requires more work for the lab instructor to load students' programs to the server manually. To achieve a remote programming function, the LabVIEW web server must be separated from the control server. Some programming languages with powerful web-based function are recommended to extend the LabVIEW web server, such as JavaScript and VB. NET.

To better address the requirements from students on virtual laboratories and improve the performance of the system, a form will be developed for students and lab instructors to evaluate the performance of the system. As laboratories are implemented in more classes with more students, the evaluation and feedback will provide more ideas about system improvement and better implementation.

CONCLUSION

Remote virtual laboratories accessed through the Internet are feasible for long-distance applications. Experiences from developing this virtual laboratory and implementing it in a Computer-integrated manufacturing course show that multiple aspects must be taken into consideration to obtain adequate performance of the online laboratory. These include the connection and communication between a web server and the physical setup (machines and processes) and the connection and communication of the web server and the Internet. In the next step in adding more systems into this laboratory, data acquisition, motion control, FieldPoint controllers, Programmable Logic Controllers (PLCs), and industrial robots will be integrated together to achieve a virtual flexible manufacturing cell that can be operated and monitored through the Internet. Technologies for system integration and web-based human-machine interfaces (HMI) need to be applied for future development. The future system will also provide an ideal research platform for studying the performance of those advanced web-based technologies in manufacturing environments, and the efficiency of system integration for improving flexible manufacturing systems.

REFERENCES

Asumadu, J. A., & Tanner, R. (2006). A service-oriented educational laboratory for electronics. *Industrial Electronics, 56*(12), 4768-4775.

Carnevale, D. (2002). Engineering accreditors struggle to set standards for online lab sessions. *The Chronicle of Higher Education*, 1-3. Retrieved 25 July, 2007, from http://chronicle.com/free/2002/02/2002020101u.htm

Chen, S. Chen, R., Ramakrishnan, V., Hu, S. Y., & Zhang, Y. (2002). *Development of remote laboratory experimentation through Internet*. CMSU online library. Retrieved from http://cyrano.cmsu.edu:2048

Cooper, M., Donnelly, A., & Ferreira, J. M. (2002). *Remote controlled experiments for teaching over the Internet: A comparison of approaches developed in the PEARL project*. The ASCILITE Conference 2002. Auckland, New Zealand. UNITEC Institution of Technology, M2D.1-M2D.9

Deniz, D. Z., Bulancak, A., & Ozcan, G. (2003). A novel approach to remote laboratories. *33rd Annual Frontiers in Education (FIE'03),* (pp. T3E8-12).

Gillet, D., Latchman, A. H., Salzmann, C., & Crisalle, O. D. (2003). Hands-on laboratory experiments in flexible and distance learning. *The Journal of Engineering Education*.

Gurocak, H. (2001). E-lab: Technology assited delivery of a laboratory course at a distance. *Proceedings of the 2001 ASEE Annual Conference*.

Irwin, G. W. (2005). Nonlinear identification and control of a turbogenerator - An on-line scheduled multiple model/controller approach. *IEEE Transactions on Energy Conversion*, *20*(1), 237–245. doi:10.1109/TEC.2004.827708

Kozono, K., Akiyama, H., & Shimomura, N. (2003). *Development of distance real laboratory system*. CMSU online library. Retrieved from http://cyrano.cmsu.edu:2048

Liou, P. S., Soelaeman, H., Leung, P., & Kang, J. (2002). *A distance learning power electronics laboratory*. CMSU online library. Retrieved from http://cyrano.cmsu.edu:2048

Salzmann, C., Latchman, H. A., Gillet, D., & Crisalle, O. D. (2003). Requirements for real-time laboratory experimentation over the Internet. *The Journal of Engineering Education*.

Saygin, C., & Kahraman, F. (2004). A web-based programmable logic controller laboratory for manufacturing engineering education. *International Journal of Advanced Manufacturing Technology*, *24*, 590–598. doi:10.1007/s00170-003-1787-7

Thamma, R., Huang, L. H., Lou, S., & Diez, C. R. (2004). Controlling robot through Internet using Java. *Journal of Information Technology*, *20*(3).

ADDITIONAL READING

Hua, J., & Ganz, A. (2003) A New Model for Remote Laboratory Education Based on Next Generation Interactive Technologies. Frontiers in Education Conference.

Hutzel, W. (2001) Creating a Virtual HVAC Laboratory for Continuing/Distance Education. International Conference on Engineering Education.

KEY TERMS AND DEFINITIONS

FieldPoint: FieldPoint is a proprietary method for interfacing devices to computers developed by National Instruments but it is very similar in principle to the concept of a fieldbus interfacing method used by many process control equipment suppliers.

LabVIEW: NI LabVIEW is the graphical development environment for creating flexible and scalable test, measurement, and control applications rapidly. It is the major programming tool in developing virtual control interfaces.

Process Control: Process control is a statistics and engineering discipline that deals with architectures, mechanisms, and algorithms for controlling the output of a specific process.

VIs: VIs are virtual instruments. In this chapter, VIs represent the graphical user interfaces programmed in LabVIEW environment for the purpose of motion control and process control.

This work was previously published in Internet Accessible Remote Laboratories: Scalable E-Learning Tools for Engineering and Science Disciplines, edited by Abul K.M. Azad, Michael E. Auer and V. Judson Harward, pp. 1-17, copyright 2012 by Engineering Science Reference (an imprint of IGI Global).

Section 4
Utilization and Application

This section discusses a variety of applications and opportunities available that can be considered by practitioners in developing viable and effective Industrial Engineering programs and processes. This section includes 14 chapters that review topics from case studies in Cyprus to best practices in Africa and ongoing research in the United States. Further chapters discuss Industrial Engineering in a variety of settings (air travel, education, gaming, etc.). Contributions included in this section provide excellent coverage of today's IT community and how research into Industrial Engineering is impacting the social fabric of our present-day global village.

Chapter 42
Using Serious Games for Collecting and Modeling Human Procurement Decisions in a Supply Chain Context

Souleiman Naciri
Laboratory for Production Management and Processes, Ecole Polytechnique Fédérale de Lausanne, Switzerland

Min-Jung Yoo
Laboratory for Production Management and Processes, Ecole Polytechnique Fédérale de Lausanne, Switzerland

Rémy Glardon
Laboratory for Production Management and Processes, Ecole Polytechnique Fédérale de Lausanne, Switzerland

ABSTRACT

Computer simulation is often used for studying specific issues in supply chains or for evaluating the impact of eligible design and calibration solutions on the performance of a company and its supply chain. In computer simulations, production facilities and planning processes are modeled in order to correctly characterize the supply chain behavior. However, very little attention has been given so far in these models to human decisions. Because human decisions are very complex and may vary across individuals or with time, they are largely neglected in traditional simulation models. This restricts the models' reliability and utility. The first thing that must be done in order to include human decisions in simulation models is to capture how people actually make decisions. This chapter presents a serious game called DecisionTack, which was specifically developed to capture the human decision-making process in operations management (the procurement process). It captures both the information the human agent consults and the decisions he or she makes.

DOI: 10.4018/978-1-4666-1945-6.ch042

INTRODUCTION

In fast-paced markets, companies try to improve product service level and quality while decreasing costs in order to get high market shares. Whereas some companies implement solutions to improve performance without prior verification, a wiser approach is to use computer simulation to evaluate the impact of potential solutions on the performance of the company and its supply chain. In these simulations, production facilities and planning processes are modeled in order to capture company behavior. However, the main drawback of this approach is that little attention is given to the human decisions that take place in this context. Human decisions are very complex, varying across individuals and even across time for a single individual. Traditional simulation models thus neglect this component. But this in turn limits the models' reliability; in fact, modeled companies often exhibit different behavior than their real-world counterparts.

The challenge is to be able to capture how human decisions are made, use this knowledge to develop reliable human decision-making models, and then implement these models in computer simulations. For this purpose, the first task is to capture human decisions as they *are* made rather than how they *should be* made. Capturing actual human decisions is not straightforward, however, because people are not very good at verbalizing what they know (Vermersch, 2006).

The utility of the conventional simulation approach in studying system behavior in general has been proved (Robinson, 2005) even though it does not involve active user participation during simulation runs. However, for the purpose of knowledge elicitation (Edwards et al., 2004) as well as user training or education, using more advanced simulation technique that integrates visual simulation and user interaction (Van der Zee & Slomp, 2009) should be a promising approach.

This chapter presents a serious game called *DecisionTack* that was specifically designed to capture the human decision-making process in a procurement context. The main motivation for developing the game is to be able to take full advantage of simulations that include active user interaction for the purpose of quantitatively analyzing decision-making behavior.

This serious game captures both the information consulted by the player and the decision he or she makes. This is done repetitively during the game, because an operational decision (procurement) is required from the player on a daily basis. This leads to the capture of a series of *decision versus consulted information pairs* that can later be used to develop human decision-making models. Subsequently, the outputs of the game are analyzed using four metrics that characterize each player's behavior in terms of data consultation and decisions.

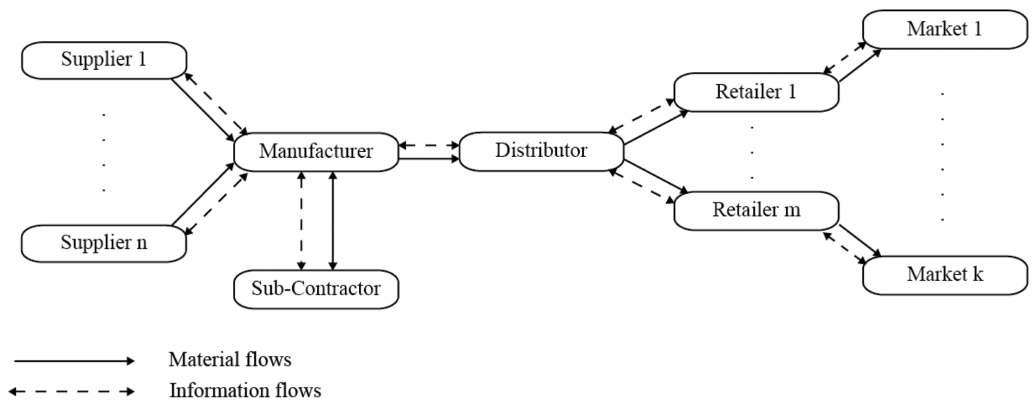
In the rest of this chapter, we lay out the basic concepts of Supply Chains, decision-making in a supply chain context, and describe previously developed serious games (*Background*); describe current weaknesses and define the goal of the research (*Motivation & Goals*); describe the details of our serious game (*DecisionTrack Game);* outline our analysis and interpretation approach (*Analysis and Interpretation*); illustrate a case study (*Application case*); discuss strengths, weaknesses and further challenges of our serious game (*Issues & Controversies*); outline potential further development (*Research Directions*) and draw final conclusions (*Conclusions*).

BACKGROUND

1. Supply Chain

In the current global economy, enterprises do not act as isolated companies, but are integrated in complex networks involving many entities (manufacturing, transportation, warehousing, etc…) that are linked by complex material flows (such as products and components) and informa-

Figure 1. Schematic representation of an enterprise network (supply chain)

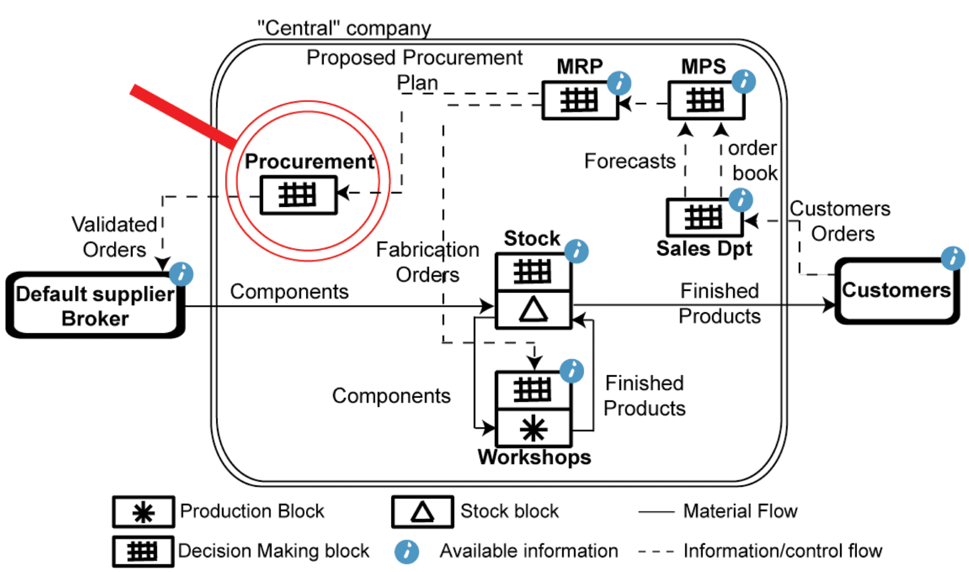


tion flows (such as customer orders or production orders). This is schematically illustrated in *Figure 1*. Terms such as 'Supply Chain' or 'Value Adding Network' are used to describe these complex networks. For simplicity, we will just use the term 'supply chain'.

Within a manufacturing company (one constitutive entity of a supply chain, see "Manufacturer" on Figure 1), the main material and information flows can be schematically described as illustrated in Figure 2.

Customer orders are received. They are entered into the order book of the company. The order book serves as the basis, together with market demand forecasts, to create a plan called Master Production Schedule (MPS). The MPS contains confirmed and expected customer orders, listed according to their desired delivery dates. Based on the MPS, a procedure is run to anticipate the need for product assembly, part production and component procurement. This procedure, called Manufacturing Resources Planning (MRP), is widely used in repetitive manufacturing.

Figure 2. Schematic representation of the material and information flows within a manufacturing company



2. Decision Making in a Supply Chain Context

In each entity of a supply chain, people are constantly making decisions; these decisions are the cornerstone of the company management and have very important and direct consequences on company performance. Human decisions can be classified according to the time horizon affected by the decision as strategic (long-term), tactical (medium-term) or operational (short-term). In supply chain management, strategic decisions are tied to the network's configuration (for example, the selection of a manufacturing location). Tactical decisions involve the calibration of the network's main management parameters (for example the level of safety stocks). Finally, operational decisions are related to the execution of repetitive tasks (for example launching production orders).

Simulations are often used in tools that support strategic and tactical decisions. If these simulation models are to help humans to make strategic and tactical decisions, they must be able to reliably reproduce the behavior of the actual Supply Chain. But the supply chain is strongly affected by operational decisions that are continuously being made by humans. Paradoxically, operational human decision-making has hardly been taken into account in Supply chain simulation models, thus limiting the reliability of these tactical and strategic decision-making tools.

Operational decisions in a supply chain context are characterized by their repetitive nature; i.e. the same decision type (for example launching a production order) must be frequently made (for example daily). The decision situation may change for each decision occurrence, however. The decision made is thus dependent on both the decision context and on the human decision-making behavior, as schematically represented in Figure 3.

Many operational decisions are made in a supply chain context, from shipping to planning and procurement. In particular, one output of the MRP procedure described above is a set of lists of time-paced propositions for launching assembly, production and procurement orders. A human decision is then required to execute the MRP-proposed orders. This decision involves confirming the MRP propositions, modifying them (date and/or quantity) or grouping some of them. In the specific case of the procurement process considered here, the operational decision can be illustrated as shown in Figure 4.

The decision elicitation problem can therefore be formalized and formulated as follows:

Each planner j makes decisions D_{ij} at time i, according to the decision context he/she perceives at the time i (DC_{ij}). This decision context

Figure 3. Schematic representation of an operational human decision making process in a supply chain context

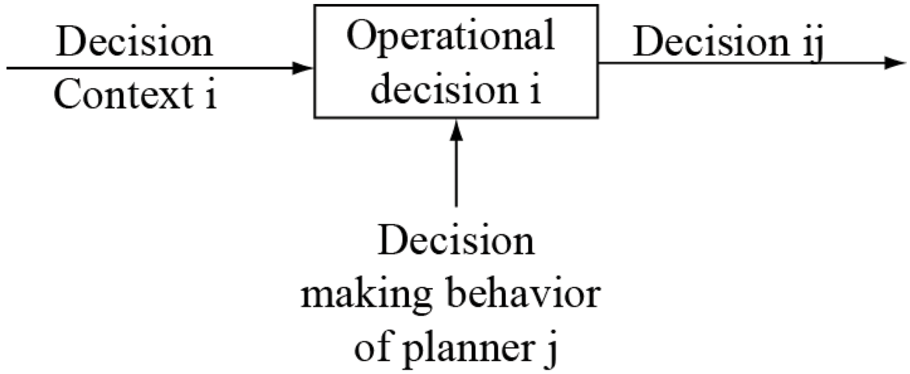
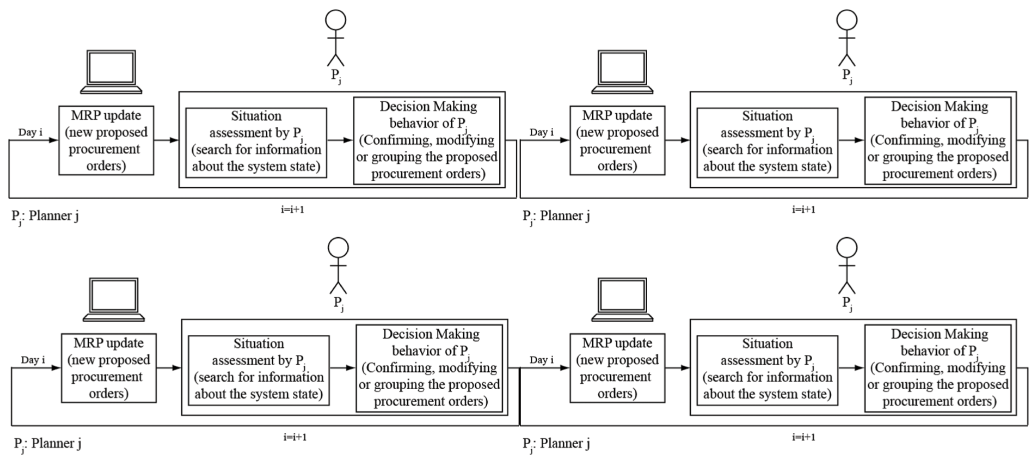


Figure 4. Illustration of the procurement decision-making process



encompasses the updated proposed procurement plan (MRP_i), as well as the collected information (CI_{ij}) gathered by the planner j at time i. Thus, the decisions D_{ij} made by the planner j at time i can be expressed as:

$$D_{ij} = f_j\ (DC_{ij}) \tag{1}$$

$$D_{ij} = f_j\ (MRP_i, CI_{ij}) \tag{2}$$

Consequently, the elicitation process consists of identifying the decision-making behavior f_j of planner j in order to predict planner j's decisions according to the information at hand. The first task in identifying f_j is to capture the decision inputs and outputs:

- Decision inputs:
  - MRP proposed orders (MRP_i), which create the decision alternatives;
  - The information collected by the planner j (CI_{ij}).
- Decision outputs:
- MRP proposals modified and validated by the planner's orders (D_{ij}).

3. Serious Games and Dynamic Decision Making

Serious games have been used for several decades for studying dynamic decision-making (DDM). DDM encompasses the decision-making processes that take place within an environment characterized by *dynamics*, *complexity*, *opaqueness* and *dynamic complexity* (Gonzalez et al., 2005). In this paper, the authors review the ten most significant serious games that have been developed since the 1980s. One of these is of particular interest here, as it involves operations and supply chain management. *Beer Game* was initially developed in the 1960s as a board game before being converted into a serious game in the late 1980s by Sterman (Sterman, 1989). The *Beer Game* has 4 entities (factory, distributor, wholesaler and retailer) that provide the market with beer. The goal is to achieve the highest service level with the lowest costs. Players are separated in four groups, each group being in charge of the replenishment decisions of a single supply chain entity. This game was popular with users (master's students and managers) because it illustrates supply chain dynamics and the virtues of collaboration across supply chain partners very well. However, one disadvantage of this game is that it is difficult to use it to understand player decision-making behavior. Indeed, the *Beer Game* does not track the information consulted by players, and thus the causal relations between the system state and the decisions made by the players cannot be explicitly rendered. This drawback is also observable in a simulation game developed to study maintenance decisions in Ford assembly lines (Robinson et al., 2005). In this game, a production process is simulated until a breakdown occurs, at which point the simulation stops and players (maintenance operators) are presented with the breakdown characteristics. Then they are asked to make a decision among a discrete set of alternatives (repair now, repair later, ask somebody else to do it, etc.). Each breakdown instance is recorded including both the breakdown characteristics and the corresponding decision. The authors then used this set of instances to model and reproduce maintenance operator decision-making behavior. However, by presenting the entire set of breakdown characteristics to maintenance operators at each breakdown, the assumption is made that all of the breakdown characteristics are equally important in the decision making process, while some of them would probably not have been consulted if the operator had had to look for relevant information by himself.

Serious games are a very efficient way of representing complex and dynamic decision making contexts (time dependency, feedback loops, endogenous and exogenous variations), but often fail to handle a very important phase of dynamic decision making - the *situation assessment* -- that can vary greatly from person to person.

MOTIVATION AND GOALS

Conventional data collection techniques such as observation, questionnaires and interviews are not well suited for the elicitation process described above (Naciri, 2010). Several hurdles (the time required, difficulty in designing suitable questionnaires, difficulty that interviewees have in explicitly describing their actions and decisions) make it difficult to rely on such techniques to collect relevant data for analyzing and modeling human decision-making behavior.

On the contrary, serious games seem to be better suited for the elicitation of human decisions in an operational context. They have several demonstrated benefits:

- First, serious games require players to "act" and not to "explain," therefore, "action satellites" can be avoided. According to Vermersch (Vermersch, 2006), "action satellites" refer to the four dimensions of the action (context, judgments, theoretical knowledge, goals) that interviewees often cite instead of discussing the action (or decision) itself;
- Second, in serious games several player actions can be recorded, giving results at a faster pace than using conventional techniques;
- Third, it is possible to analyze how decisions made at time t influence the environment at time t+1, making it possible to capture the dynamic aspects of decisions and, in particular, the notion of feedback;
- Finally, serious games create an environment in which player actions can be recorded, as well as the context in which decisions occur. Thus, once the simulation is finished, it is possible to pair the simulation context with the decisions that were made.

Participatory simulation via serious games thus appears to be a well-adapted tool for capturing the various dimensions of human decision-making, and thus for obtaining a quantitative understanding of human decision making behaviors.

The goal of this work is to develop a serious game that will elicit operational human decisions in a supply chain context, more specifically, in procurement.

Because the objective is to reliably generate human decision-making situations that are representative of what happens in the industry, the game must fulfill the following criteria:

- Interfaces similar to typical industrial tools: this ensures that players (procurement agents) act (make decisions) as they would in their working environments. It thus ensures that the information collected in the virtual environment is representative of procurement agent behavior in the "real world."
- Available information similar to that in actual ERP systems: this ensures that players are familiar with the information at hand in the simulator and that no specific training is required to explain how the information is displayed.
- Decision pace to avoid player stress: one of the main drawbacks of real-time serious games is that the pace is often accelerated, giving the player little time to make decisions. Consequently, players must control the simulation pace in order to get enough time for making relevant decisions.

The main function of the developed serious game, *DecisionTrack,* is to collect data to identify:

1. What information procurement agents are interested in, and
2. What kinds of decisions procurement agents make.

According to Van der Zee & Slomp (2009), a framework for game design has four phases:

- Initialization: definition of the scope and objectives of the game.
- Design: Detailed development of the basic ideas formulated in the initialization phase. The outcome of this phase is a simulation game concept.
- Construction: construction of the game using software or other physical elements.
- Operation or game running: actual use of the game, which may include a test of the game for its intended purpose.

We have covered the "Initialization" phase of the game design framework in this section. The next section explains development methodology of the *DecisionTrack* serious game.

DECISIONTRACK GAME: FROM DESIGN TO IMPLEMENTATION

1. Definition of Game Concept

The *DecisionTrack* game was designed according to the following objectives:

1. To provide the player with a realistic decision making context (similar to the one he/she is usually involved with),
2. To allow the player to consult the information he/she feels relevant,
3. To allow the player to make the decisions he/she feels relevant,
4. To capture player actions (consulted information and decisions) and to save this information as readable data (log file).

The underlying motivation behind these four objectives is to build a decision making context that is similar to the one players are used to working with, in order to capture decision making situations that are as close as possible to those they encounter in reality. This can be done not only by creating realistic decision-making situations, but also by not constraining players to limited sets of data or decisions.

Finally, keeping track of player's actions makes it possible to subsequently identify the relationship between the decision-making context and the decision made.

2. Virtual Decision Making Environment

The virtual decision making environment chosen for this game is a two-tier supply chain, in which the player is in charge of the procurement process. His/her task is to modify (if needed) and validate the procurement orders based on MRP propositions, as described in the previous section. In order to provide players with a familiar decision making environment, a virtual supply chain with a commonly used production management policy (make to stock) is used. It is illustrated in Figure 2.

Several entities representing departments of the company such as *production*, *warehousing*, *planning* (i.e., *MPS*, *MRP* and the *procurement* process under study) are included in this virtual environment. It also includes some external entities such as *customers* and *suppliers*.

The circles (containing the letter "i") on the top of entity icons in Figure 2 indicate that information concerning the corresponding entity is available for player consultation.

Table 1 summarizes the game elements that are included in *DecisionTrack*.

The following subsections describe in detail how the game elements are constructed.

3. DecisionTrack interfaces

The game is developed in Java 1.6 (Java SE 6) fully benefiting from the built-in Java Swing libraries in order to implement complex user interactions. The main goal of *DecisionTrack* interfaces is to provide information about the company and its supply chain. This information, which is modified daily, enables players to update their knowledge of the system.

Displayed Information

The relevant information to be displayed was identified by conducting an analysis of the procurement process with experts in the procurement field. This analysis provided valuable insight into the kinds of information procurement agents consult, and the kinds of decisions they make.

Two ERPs (Enterprise Resource Planning) widely used in Switzerland, SAP (SAP, 2011)

Table 1. Summary of DecisionTrack main game elements

Game elements	Definition
Model and Scenarios	- A 2-tier supply chain - Context: a manufacturing company, MRP information to consult - Decision making options: accept the MRP propositions, postpone, anticipate, group orders by modifying the order launching date
Game process	1. Presentation of the game to each player 2. Each player plays the game 3. Data collection and analysis 4. Player decision making modeling
Events	New customer orders, new forecasts, component deliveries from suppliers (on time and late deliveries)
Periods	One period per simulation day A complete run with a single player lasts at least 30 periods
Roles	Procurement agent
Results	Performance indicators such as inventory levels and service level
Indicators	MRP data, supplier-related data, customer-related data Above-mentioned performance indicators (see "Results")
Symbols, Materials	Various user interfaces (windows) that mimic real ERP systems in companies

and Proconcept ERP (Proconcept, 2011) were investigated, in order to design *DecisionTrack*'s interfaces to mimic the interfaces procurement agents are accustomed to working with.

Tracking Methods

As stated before, players can navigate through *DecisionTrack* interfaces in order to update their knowledge of the decision context. Because different people search for information in different ways, it is essential to track which specific information is consulted. This is accomplished by isolating the information related to each supply chain entity in a separate tab of the game window. Each tab is associated with a mouse-listener that is activated once the tab is selected. The record of the activated tab is stored in a "log file".

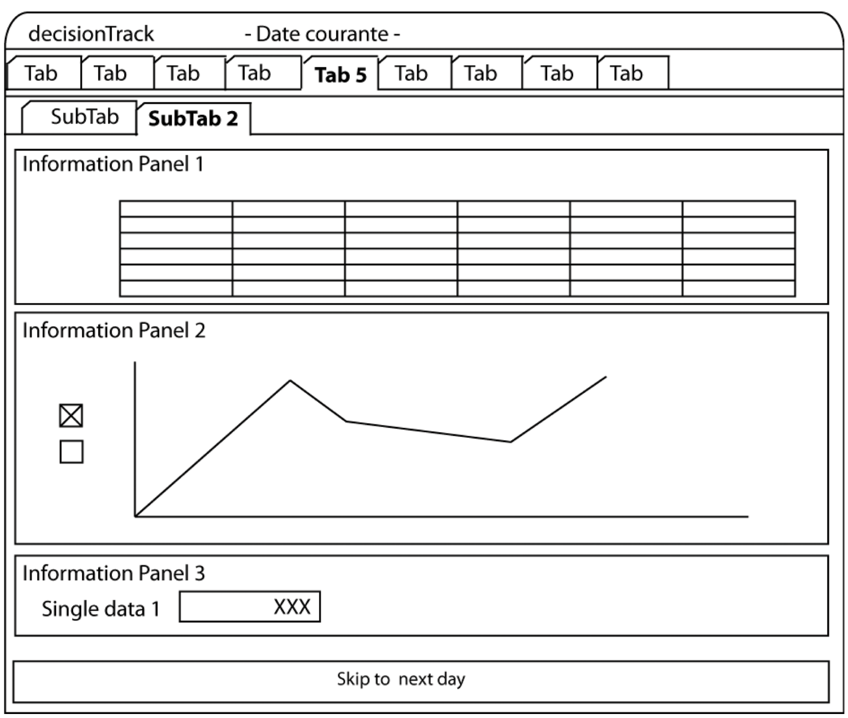
In this way it is possible to track which supply chain entity a players is interested in. Because several pieces of information may be needed to describe a supply chain entity, a single interface may not have enough space to correctly display the whole set of information. In these cases, tabs may contain two or more sub-tabs between which the entity-related information is split.

In the case where a single sub-tab contains heterogeneous information, checkboxes are added to help track the consulted information. These checkboxes are by default unchecked, which makes the corresponding information unavailable. When a player is interested in a piece of information, he/she checks it, releasing the corresponding information.

In sub-tabs and their checkboxes, the technique of tracking consultations is similar to the one described for the tabs. Using mouse-listeners which are attached to each graphic element, the name and the value (when relevant) of the consulted information is recorded in the log file.

Tabs, sub-tabs, information panels and checkboxes are organized in a hierarchical way as illustrated in Figure 5.

Figure 5. Illustration of the hierarchical structure of the information within a DecisionTrack window



Interfaces for Collecting Data

The central part of Figure 6 illustrates the window corresponding to the *Customers* tab. This interface provides various pieces of information. The first information panel (1) is the evolution of the service level (ratio of the orders delivered on time) since the beginning of the simulation. The second information panel (2) contains only the current service level, and the third panel (3) contains the list of pending orders, (orders that have not yet been delivered).

To get to this specific interface, the player must go through the following steps:

1. Select the *Customers* tab
2. Select the *Service level* sub-tab
3. Activate the *Service level* checkbox in the upper information panel
4. Activate the *Service level* checkbox in the second information panel

The right hand side of Figure 6 shows the corresponding data recorded in the log file.

Interfaces for Making Decisions

All the interfaces except *Purchaser* contain information for consultation that cannot be directly modified by the player. The information contained in the *Purchaser* tab (proposed procurement orders generated by the MRP algorithm) can be updated by the player with or without modifications.

The *Purchaser* interface is illustrated in Figure 7. As shown in the bottom panel, several procurement orders with a proposed launching date and quantity are displayed. According to his/her knowledge of the system state, the player can modify the propositions by clicking on an order, and then on the line corresponding to the new launching date. Four kinds of modifications can be made:

- *Anticipation* (the new order date is closer to the current simulation date than the former one),
- *Postponement*,
- *Grouping* (when the new launching date of the order corresponds with the launching date of existing orders),
- *Do Nothing* (no modification is made to the proposed order).

Once the player has chosen one of these alternatives, he/she can click on the padlock to validate the order and send it to the supplier.

All the decisions made by the player are recorded in the log file, so that the correspondence

Figure 6. Illustration of a specific interface (left) and the corresponding log file records (right)

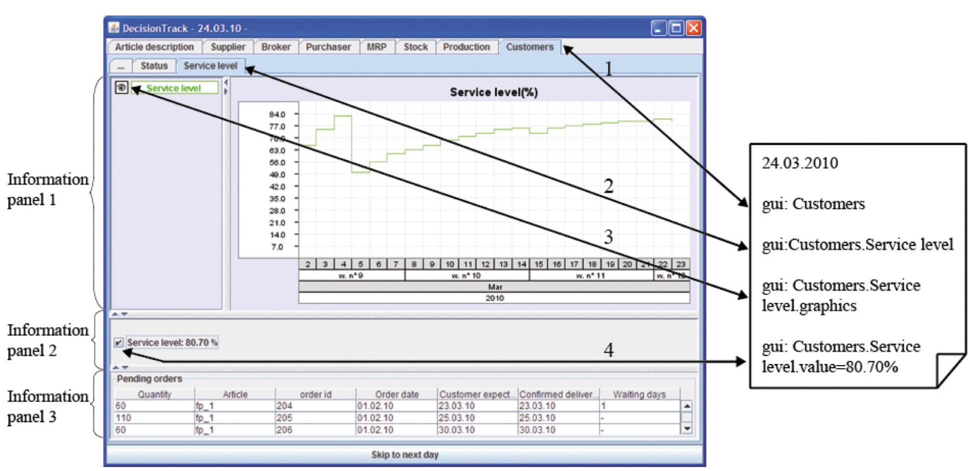
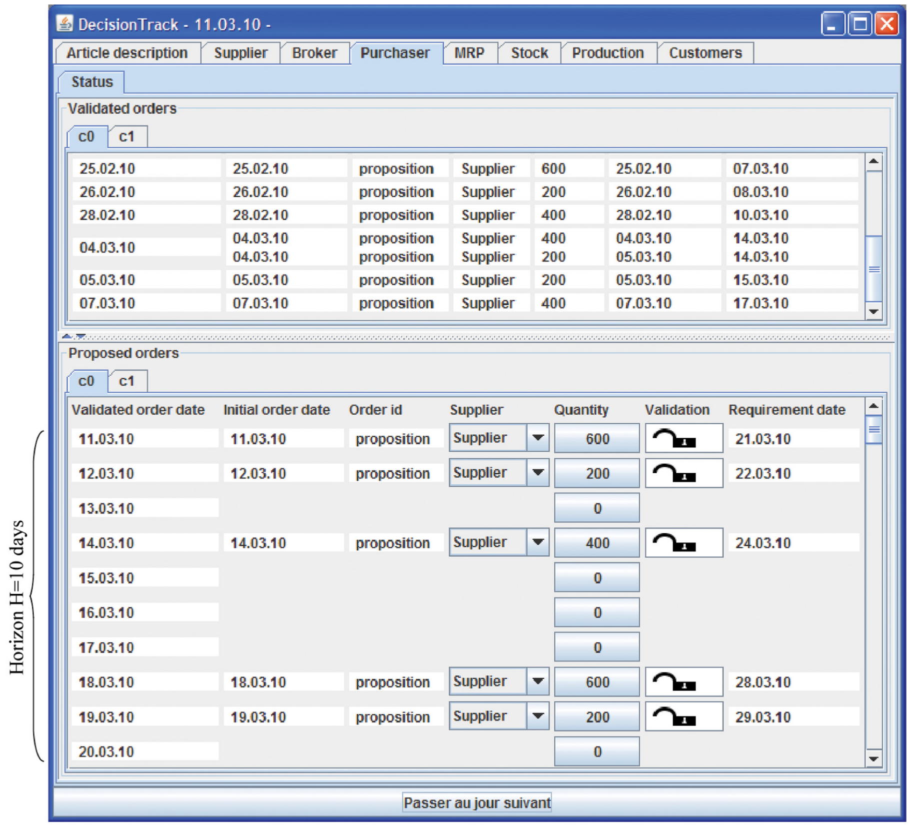


Figure 7. Snapshot of the "purchaser" tab



between player knowledge of the system state and the decisions he/she has made can be identified.

4. DecisionTrack Scenario

Game scenarios refer to the evolution of game variables across time. When designing a new game, two categories of variables must be differentiated. The first category encompasses variables that cannot be modified by the player -- *exogenous* variables. The second category relates to variables that can be modified by the player through his/her decisions -- *monitoring* variables. The latter help the player monitor the impact of his/her decisions on company performance.

The *exogenous* variables are identified through a literature review and discussions with domain experts. These variables must vary across time to create decision-making situations that require specific attention.

In this research, the goal is to identify how procurement agents make decisions according to the evolution of the environment (suppliers and customers). Table 2 provides the list of all *exogenous* and *monitoring* variables used in this study. Among the *monitoring* variables, it is worth noting that the service level value depends on the way the player updates the proposed procurement plan.

The selected *exogenous* variables that should vary across the simulation are those that are related to the supply chain environment, such as component delivery times, market behavior (customer orders), and forecasts (see Table 2).

By making the above-mentioned *exogenous* variables change across time in the game scenario, a difference appears between planning algorithm

Table 2. Exogenous and monitoring variables

Displayed information	Location	Exogenous variable	Monitoring variable
Customer orders and forecasts	- *Customers* tab - *Status* sub-tab	YES	NO
Actual and predicted component delivery time	- *Supplier* tab - *Graphics* sub-tab	YES	NO
MRP tables	- *MRP tab* - *Status* sub-tab	NO	YES
Inventory levels	- *Stock* tab - *Status* sub-tab - *Graphics* sub-tab	NO	YES
Work in process	- *Production* tab - *Status* sub-tab	NO	YES
Service level	- *Customers tab* - *Service Level* sub-tab	NO	YES

propositions (that are based on MRP theoretical parameters), and the current simulated situation. Such gaps create critical decision-making situations that require the player to make decisions that impact the company performance. The gaps between the planning algorithm's recommendations and the evolution of company's environment correspond to realistic and very common situations that routinely appear in companies in which the MRP parameters are not updated according to variations in the environment.

Supplier Delivery Time Variations

Supplier delivery times are set in the scenario to differ from the theoretical delivery time introduced in the MRP algorithm. In this way, actual delivery time may be either 1) longer than the theoretical one (which leads to delivery delays), or 2) shorter than the theoretical one (which leads to deliveries that enter the stock earlier than expected). The scenario is set so that situation 1) occurs most often, encouraging the player to make decisions instead of choosing the status-quo alternative.

Supplier-predicted delivery time data are provided in the form of a table in the *Supplier* tab.

In addition to the predicted delivery time, the actual delivery time is reported after each order delivery. All three delivery times (theoretical, predicted and actual) are reported as shown in Figure 8.

Forecast Patterns

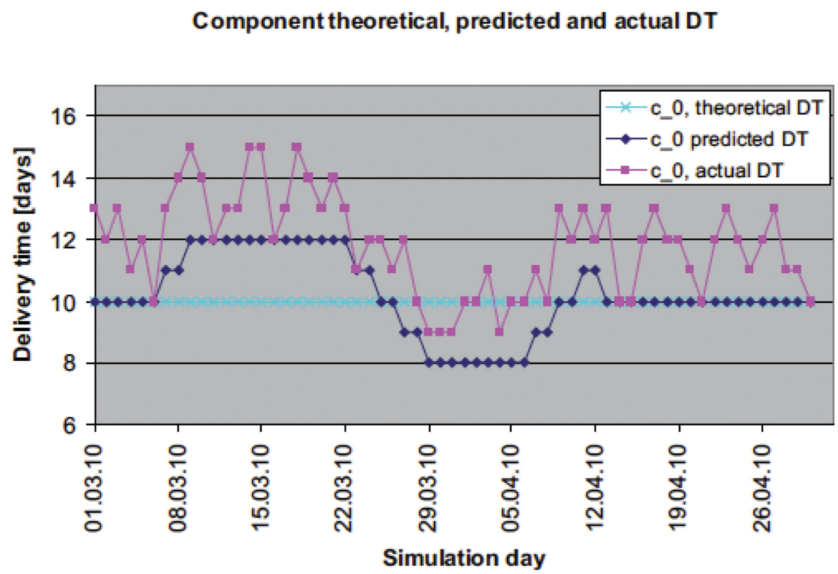
In addition to customer orders that represent the actual impact of the market on the "central" company, forecasts are designed to anticipate market requirements. The remainder of this section discusses forecast patterns (evolution of the forecasted demand over the planning periods) and whether they adequately predict customer order patterns.

Forecasted demand and customer order patterns are designed so as to lead to either overestimation (Delta>0) or underestimation (Delta<0) of the future demand. These two situations correspond to the Case 1, Delta>0, and to the Case 2, Delta<0, in Figure 9.

In addition, both patterns vary across time. Figure 9 shows an increase in both forecasts and customer orders during the end of April, followed by a decrease and stabilization by the end of May.

There are two arguments for introducing relative and absolute variations of customer orders and forecasts across time. First, it attempts to capture the real market characteristics of *seasonality* (for example, the increase of customer orders during the month of May) and *volatility* (modeled by the inability of forecasters to correctly predict market behavior). Second, it creates different decision-making situations (under- and over-estimation of actual demand), making it possible to record player reactions in several decision contexts and capture

Figure 8. Evolution of forecasts and customer orders for product fp_0 (DT stands for Supplier Delivery Time)



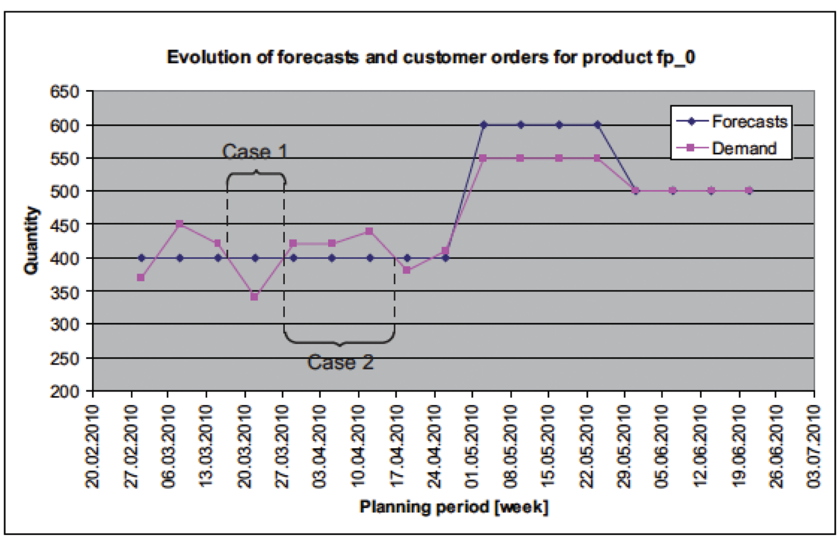
the variability that may exist among players in terms of decision-making behavior.

Because the production process is downstream of the procurement process, there is no need to make changes in the production lead-time. It is thus fixed to 8 days throughout the simulation.

Assigned Goals

Players are asked to validate the procurement plan proposed by MRP on a daily basis and to achieve the following two objectives:

Figure 9. Variations between forecasts and demand as set in the game scenario



- Minimize inventory (inventory levels are available in the "Stock" tab),
- Maximize service level (service level is available in the "Customer" tab).

These two objectives are justified by the fact that the two most common assignments for procurement agents are to control inventory and service levels:

- **Inventory levels:** high inventory levels usually correspond to high amounts of immobilized capital and may represent a risk in case of item obsolescence. Moreover, high inventory levels require large storage areas and may induce high costs in terms of temperature and hygrometry control. Procurement agents are thus often encouraged to minimize their stock, unless this leads to shortages.
- **Service level:** this performance indicator represents the ratio of the orders delivered on time by procurement to production (production lead-time fixed to 8 days). It thus provides relevant information about the extent to which the procurement agents fulfill their mission of assuring a reliable supply to the production department.

5. Simulation Sequence

Experimental Setup

In order to collect data in a consistent way across participants, an experimental procedure is designed that provides all players with the same knowledge of the game. It consists of the following four steps:

1. The *DecisionTrack* game and goal assignment is presented, and a slide show explains the displayed information in the game graphical user interface. Duration: 10-15 min.
2. Each players browses the *DecisionTrack* interfaces and the displayed data in each tab and sub-tab is described to the player. Duration 10-15 min.
3. Each player plays the *DecisionTrack* game for a minimum duration of 30 simulation days. Duration: around 1 hour.
4. Each player is interviewed to verifying his/her understanding of the game, as well as the strategies he/she engaged during the game. Duration: 20 minutes.

Simulation Day Description

To provide players with enough time to consult data and make decisions, the length of the simulation day is not constrained by the game, but dictated by the player. The player decides when to skip to the next day by clicking the *Skip to next day* button in the bottom of the window. This updates all the game variables, in particular, the procurement plan proposed by the MRP, which is updated by taking into account former decisions, current inventory levels, and variations in demand.

As described in step 3) of the experimental procedure, the average duration of a game is around one hour. It may vary around this value depending on the amount of time players spend on each simulation day, which in turn depends on the time each player needs for assessing the system state and for making his/her decisions.

ANALYSIS AND INTERPRETATION

The analysis of *DecisionTrack* outputs is based on the definition of appropriate metrics, which allows each player's behavior to be characterized in a quantitative fashion. This section discusses how to characterize a player's decision-making behavior after his/her game play. (Notation: Each experiment is labeled as Xi, where X refers to the company id and i indicates the number of the participant. For example, A2, refers to the game played by player number 2 of company “A”.)

1. Result Characterization Using Metrics

This section presents four metrics that were specifically developed to characterize player behavior in a procurement context. Two of them (M1 and M2) concern the search for information within the game interfaces, while the other two (M3 and M4) characterize the decisions made by the players. Each of these metrics is associated to a graphical illustration that helps to visually understand how the player’s behavior varies across participants.

1. Metric M1

Metric M1 is the average number of tabs consulted on a daily basis by the player over the game. The objective is to measure the player’s willingness to update his/her knowledge of the system.

2. Metric M2

Metric M2 is the number of tabs representing at least 80% of the total number of consulted tabs over the game. M2 provides cues about the extent of the information the player bases his/her decisions on.

3. Metric M3

Metric M3 is the maximum number of days between an initial order launching date and its new launching date after the player’s decision. Thus, it measures the amplitude by which the player modifies the initial procurement plan.

4. Metric M4

Metric M4 is the most distant day, from the current game day, for which a decision is made. This metric captures how early a player modifies the proposed orders, either by validating, anticipating, grouping or postponing them. Thus, this metric indicates to what extent the player focuses his/her attention on a short or a long-term horizon.

2. Clustering of Decision Making Patterns

Each metric is accompanied by a corresponding graphic that offers a visual feedback about a player’s behavior. In addition to this visual feedback, the metrics quantify an intangible process (tabs and number of times they are consulted, decisions that vary in terms of nature and amplitude).

This quantification lets us pursue the analysis further, by identifying clusters of players who behave in the same way according to the four metrics.

The clustering process is run using a k-means algorithm. Because the number of clusters is considered as a parameter (Ray & Turi, 1999), several clustering iterations are carried out with various numbers of clusters. The number of clusters is then set by selecting the iteration with the lowest error, based on the Square Mean Error and the Davies-Bouldin index (Davies & Bouldin, 1979).

APPLICATION CASE

In this section, a procurement decision application of the *DecisionTrack* game is described. For this experimentation, 14 games were played by procurement agents. Then the simulations results were analyzed according to the previously described metrics and clustering algorithm.

1. Player Characteristics

Players were purposely chosen among professionals and not students. Because of the focus on the procurement task, procurement agents were selected for participating in the game, based on these three constraints:

- **Company management policy**: We restricted participants to those whose companies use a Make to Stock policy using MRP planning. This ensures that players know how the MRP procedure works, which in turn facilitates their understanding of the game.
- **Company size**: We limited participants to those working in medium-sized and multinational companies. This ensures that people in charge of procurement are dedicated to this function, which may not be the case in small companies.
- **Geographical location**: for practical reasons, players were selected among companies located within a two hour driving distance.

2. Analysis of the Experimental Results

After experimental game runs with several players were concluded, the simulation outputs were analyzed to characterize each player's decision-making behavior. The analysis is based on the four above-mentioned metrics, which allow a quantitative characterization of the player's decision-making behavior. This is complemented by graphical representations that provide visual explanations of how a player behaves during the game.

Metric M1: Average Number of Tabs

The two graphs in Figure 10 show a high variability in data consultation patterns; player F1 regularly updates his knowledge of the decision context (M1=3.34) by consulting the available tabs, while player C3 does it on a less regular basis (M1=1.85).

Metric M2: Most Often Consulted Tabs

The illustration in Figure 11 provides the consultation intensity per tab (see number of tab-consultations close to the name of each tab) and allows the identification of the most consulted tabs. For instance, player H1 in Figure 11a shows a particular interest for the *MRP*, the *Stock*, the *Production* and the *Supplier* tab (M2=4). In Figure 11b, player H3 shows interest in a limited set of tabs (M2=2), i.e. the *MRP* and the *Stock* tabs.

This data reinforces the hypothesis that different players do not necessarily base their decisions on the same set of information, and they do not update their knowledge of the decision context at the same pace.

Figure 10. Graphical illustration of the metric M1 for players F1 and C3

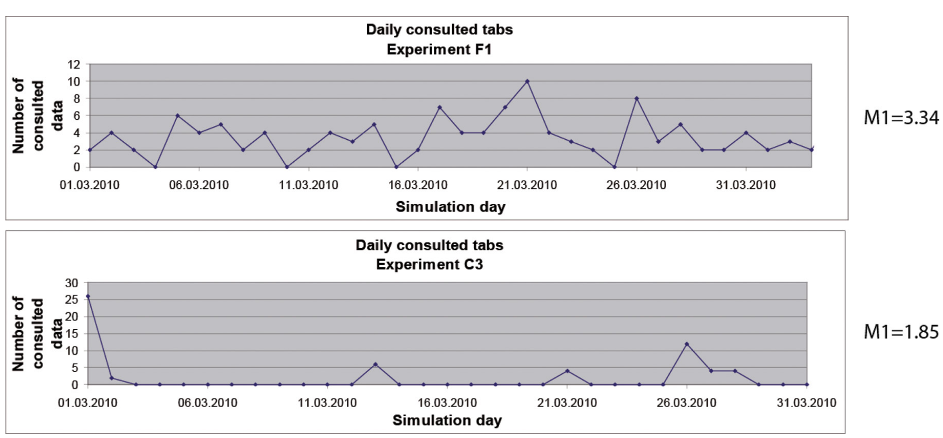
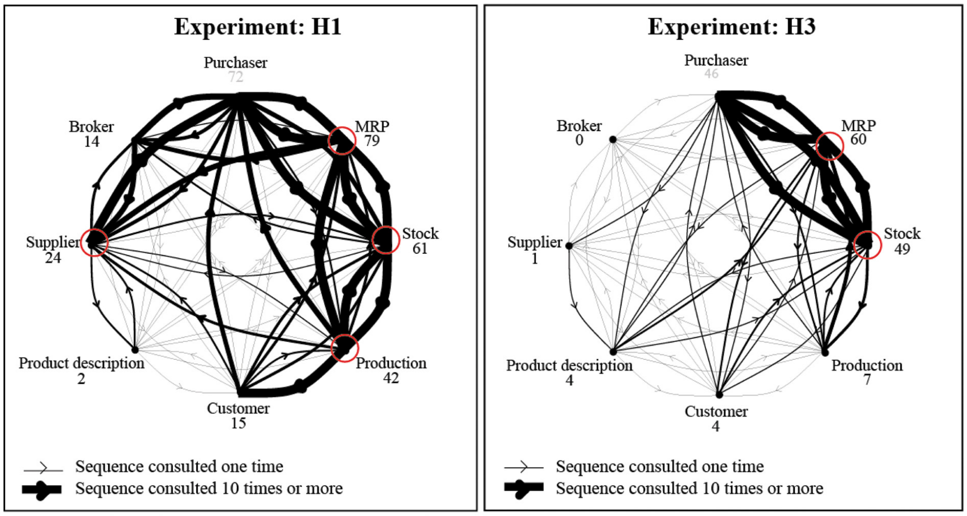


Figure 11. Graphical illustration of the metric M1 for players F1 and C3



M2=4 M2=2

Metric M3: The Difference between the Initial Setting and the Modified Date

Figure 12 shows two different player behaviors. Player D2 (M3=1) seems to be very confident in MRP propositions as he/she always (with two exceptions) validates the MRP propositions. Player D1 (M3=9), however, often anticipates the orders, which may mean that his/her knowledge of the company and the supply chain (M1=3.375 and M2=4) encourages him/her not to blindly follow MRP recommendations.

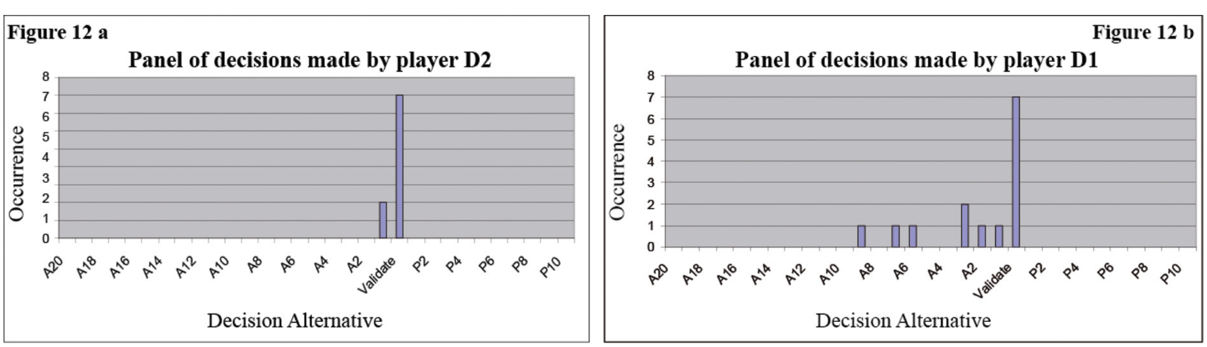
Metric M4: How Early Was the Decision Made?

In Figure 13a, we see that player A2 takes a long horizon into account as his/her decisions are contained in a 19-day span. Player C2 (Figure 13b) makes decisions on a short-term horizon -- M4=4.

Clustering

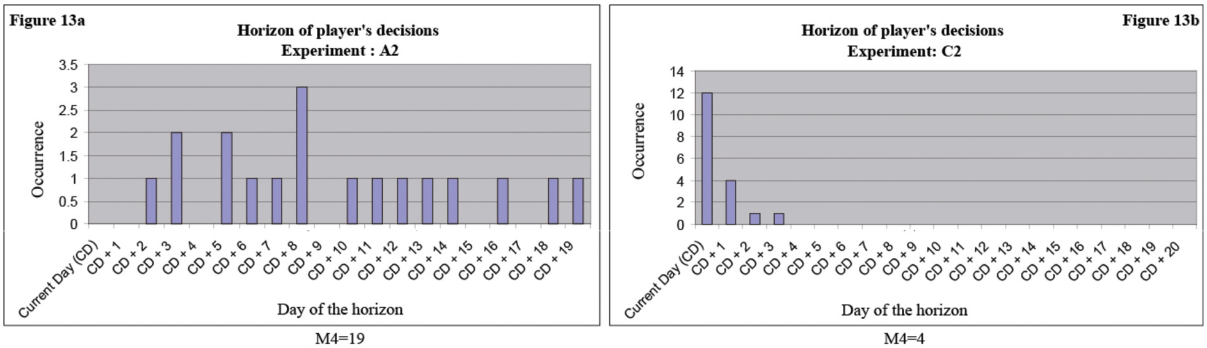
Once the fourteen experiments (players A2, B1, C2, C3, D1, D2, E1, F1, F2, F4, G2, H1, H2 and H3) are completed, the four metrics are com-

Figure 12. Graphical representation of the metric M3 for players D2 and D1



M3=1 M3=9

Figure 13. Graphical illustration of the metric M42 for players A2 and C2



puted for each experiment using the information recorded in the log file. Then, each experiment's metrics are ordered in a four-element vector, and the goal of the clustering process is to identify whether similarities exist among the 14 vectors, indicating similarities in player behavior.

The clustering results obtained using the k-means algorithm are outlined in Table 3.

Table 3 shows that the constituted clusters do not contain players who are hired by the same company.

This is evidence that the variations across players are not only related to the players' specific working contexts, but also to individual personality characteristics (such as risk-seeking and risk-averse behavior).

This is confirmed by Figure 11 and Figure 12, which show that players from the same company exhibit completely different behavior.

3. Discussion of the Application Case

There are two main observations to be drawn from the experiments. First, players vary widely in terms of the data they consult. Some focus on a limited range of information, while others deeply investigate the decision context. This corresponds to individual differences in data consultation approaches; either a systematic consultation of the same limited set of tabs, or a problem solving approach in which the search for information is adapted to previously consulted tabs. Second, players vary widely in terms of the decisions they make. Some trust the planning algorithm and modify the MRP propositions very little, while others significantly modify the initial procurement plans.

The clustering approach allows us to identify categories of players who behave the same way. This classification can be very useful as it may permit the identification of a specific category of players that deserve further attention. For instance, cluster 4 (M1=3.65, M2=2.67, M3= 2.33 and M4=4.67) deserves less attention than cluster 1 (M1=1.69, M2=3.5, M3=18.5 and M4=19.5), because in cluster 1, the modification of the proposed procurement plan is limited, decreasing the impact of the human decisions on the company and its supply chain performance.

Table 3. Player behavior clustering analysis

Cluster number	1	2	3	4	5
Cluster components	C3, G2	D1, E1, F1, F3, H1	B1, H3	C2, D2, F2	A2, H2

ISSUES AND CONTROVERSIES

The several experimental tests carried out before the final data collection tests allowed us to identify several hurdles related to the use of serious games. They are detailed below, along with measures we adopted to overcome them, as far as was possible.

1. Player Behavior Distortion

When using laboratory experiments in general and serious games in particular, there is a risk that the observed player behavior is distorted in comparison with his/her actual behavior in the "real world." This gap can be explained by three factors: *stakes, social norms* and *scrutiny* (Levitt et al. 2007).

- *Stakes* within games often differ from the stakes in real world. For example, the consequences of badly managing the procurement orders in *DecisionTrack* are less critical than those engendered by badly managing the procurement orders in a real company (overstocking, shortage, customer dissatisfaction, etc.). Because this, players may show a lack of commitment in a virtual game scenario. In our experiments, *stakes* have a limited impact on player commitment, as all the players are professional procurement agents who participate in the name of their company, thus who have implicit motivation to behave in the best way to uphold the good image of their company.
- *Social norms* refer to the fact that laboratory experiments lack the "rich real-life" context that exists outside the laboratory. Because this aspect typically characterizes "social experiments", it does not affect the consistency of the data collected with *DecisionTrack*. It is thus not discussed further.
- Finally, *scrutiny* refers to the fact that within laboratory experiments, player may act such a way to please the experimenter. This is also called the experimenter *demand effect* (Orne, 1962). To avoid this, the instructions provided before the data collection phase reiterate the researchers' interest in collecting and reproducing their decision-making behavior as it is, regardless of the quality of their decisions or the corresponding performance. Thus, players do not have any preconceived idea about particular behavior the experimenter is seeking to observe. Thus, players cannot please the experimenter by adopting a particular behavior.

2. Platform Understanding

The main drawback of providing players with a realistic, though virtual, environment is the multiplication of the data at their disposal in the *DecisionTrack* interfaces. As a consequence, players may find it difficult to handle the serious game, and to understand the interface contents in an exhaustive way.

In order to help players to correctly handle the tool, a two-phase protocol – learning phase and verification phase -- has been designed to ensure player understanding of the game. The learning phase is divided into two sub-phases that occur before the game is played; in the first, the experimenter presents a slide show describing the research approach, the assigned goals, and shows interfaces screen shots. In the second sub-phase, the experimenter browses *DecisionTrack* interfaces with the player and explicitly describes the information they contain. During the learning phase, the experimenter also answers player questions.

The verification phase takes place after the data collection phase is carried out, using an interview. The goal of this interview is to ask the player about his or her strategy and the kind of decisions he/

she has made. This verification interview is later used to compare what the player has said about his/her decision-making behavior with what has been recorded in the log file.

This verification allows us to withdraw experiments in which players did not understand the meaning of the displayed information, or in which players did not manage to log their decisions in the game because they misunderstood the game commands.

3. Identification of the Consulted Information

Although measures are put in place to track each player's search for information (tabs, sub-tabs, checkbox, mouse listeners), situations come up in which accurate identification of the consulted information remains tricky -- particularly in the case of graphs and tables. Indeed, even if the tracking tool identifies that a table is consulted, it is difficult to identify which rows or columns are actually read by players. In a similar way, it is difficult to identify exactly how a player consults a graph (trend, horizon, scales, etc.). This issue could be addressed using gaze-tracking technology, and will be discussed in the next section.

Several things may hinder the relevance of the collected data. Some are related to human characteristics and behavior while others are due to the game complexity resulting from the game realism.

However, the effort taken to run the experiments according to a validated protocol, the attention given to introduce the game and to train players, the care in the choice of words in order not to influence players, the attention given to isolate the information within the interfaces make it possible to overcome the hurdles that commonly stand in behavioral researcher's way give us a very high level of confidence in the relevance of the collected data.

FUTURE RESEARCH DIRECTIONS

To address the consulted information identification issue mentioned in the previous section, gaze-tracking devices could be used.

Gaze tracking technology does exactly what it advertises – it tracks the direction in which a user's eyes are looking. Several gaze tracking technologies exist; they vary according to the positioning of their sensors or cameras, and how they are calibrated. Conventional gaze tracking devices rely on sensors that are placed on the computer display and must be calibrated prior to the gaze direction collection phase. A more innovative approach (Noris et al., 2010) is portable devices that can capture gaze directions in non-laboratory settings (outdoor environments, workplaces). The main obstacle in making these devices applicable for the identification of consulted information within serious games is the synchronization between the eye tracking device and the information contained in the game interfaces. If this could be overcome, it would be possible to know how many rows a player is looking at in a table at a given time and the data displayed in these rows, for example.

CONCLUSION

This study shows that when consistent efforts are made to create realistic virtual environments, serious games provide a relevant approach for capturing consistent human operational decisions, while avoiding the drawbacks (e.g. verbalization, theorization) that are met using conventional data collection tools (e.g. interviews, questionnaires).

A large panel of information was gathered from experts in the field under study in order to identify the set of information players may consult and the set of decision alternatives players may make. Interfaces were organized in a way that was familiar to players, to make them comfortable with the game. In addition, scenarios were designed to introduce perturbations and to create decision situations that required players to make decisions. And before games were played, a protocol was designed to provide all players with exhaustive information about the game interfaces, the assigned goals and the commands for interacting with the game.

The present work focuses on procurement decisions, but this serious game approach could be applied to almost all other operational decisions found in Supply Chain management.

The analysis of the collected data using four metrics reveals a large variability among player behavior, which supports our initial observation that people make decisions in highly individual ways, and that this variability should be taken into account in computer simulations. This is also confirmed by the clustering analysis that shows that even people working within the same company vary in their decision-making behavior.

Future challenges for serious games include incorporating new tracking technology capabilities (synchronized gaze tracking) that will enable more accurate data collection, improving our understanding of how people make decisions according to the information they have at hand and their perception of their working environment.

ACKNOWLEDGMENT

Our heartfelt thanks go to Olivier Gobet for his outstanding participation in the development of the DecisionTrack game, as well as Mary Parlange who reviewed the English phrasing of this paper.

REFERENCES

Davies, D. L., & Bouldin, D. W. (1979). Cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(2), 224–227. doi:10.1109/TPAMI.1979.4766909

Edwards, J. S., Alifantis, T., Hurrion, R. D., Ladbrook, J., Robinson, S., & Waller, A. (2004). Using a simulation model for knowledge elicitation and knowledge management. *Simulation Modelling Practice and Theory*, *12*, 527–540. doi:10.1016/j.simpat.2003.09.003

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286. doi:10.1016/j.chb.2004.02.014

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, *21*(2), 153–174. doi:10.1257/jep.21.2.153

Naciri, S. (2010). *Elicitation and modeling of operational human decision making in a supply chain context: A study based on participatory simulation.* Ecole Polytechnique Fédérale de Lausanne (EPFL), PhD Thesis no 4725 (2010). EPFL.

Noris, B., Keller, J.-B., & Billard, A. (2010). A wearable gaze tracking system for children in unconstrained environments. *Computer Vision and Image Understanding*, (November): 1–27.

Orne, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *The American Psychologist*, *17*, 776–783. doi:10.1037/h0043424

Proconcept. (2011). *Proconcept ERP*. Retrieved from http://www.solvaxis.com/fr/Produits/ProConcept-ERP.html

Ray, S., & Turi, R. (1999). *Determination of number of clusters in K-means clustering and application in colour image segmentation*. Narosa Publishing House.

Robinson, S. (2005). Discrete-event simulation: from the pioneers to the present, what next? *The Journal of the Operational Research Society*, *56*, 619–629. doi:10.1057/palgrave.jors.2601864

Robinson, S., Alifantis, T., Edwards, J. S., Ladbrook, J., & Waller, A. (2005). Knowledge-based improvement: Simulation and artificial intelligence for identifying and improving human decision-making in an operations system. *The Journal of the Operational Research Society*, *56*(8), 912–921. doi:10.1057/palgrave.jors.2601915

SAP. (2011). *SAP ERP operations*. Retrieved from http://www.sap.com/solutions/business-suite/erp/operations/index.epx

Sterman, J. D. (1989). Misperceptions of feedback in dynamic decision-making. *Organizational Behavior and Human Decision Processes*, *43*(3), 301–335. doi:10.1016/0749-5978(89)90041-1

Van der Zee, D., & Slomp, J. (2009). Simulation as a tool for gaming and training in operations management-A case study. *Journal of Simulation*, *3*, 17–28. doi:10.1057/jos.2008.8

Vermersch, P. (2006). *L'entretien d'explicitation*. ESF Editeur.

KEY TERMS AND DEFINITIONS

Elicitation: Is a process used to help experts to share their knowledge. It consists of using several methods that make expert's tacit knowledge more explicit, and thus easier to analyze, model and reuse.

Mouse-Listener: A mouse-listener is a JAVA listener interface that records the mouse events (press, release, click, etc.) on a component (button, checkbox, etc.). Thus, mouse-listeners are either used to record user actions (for example, enters the consulted information and the decisions in the log file) or to react to user actions (for example, when the user checks the "skip to the next day" button in DecisionTrack's window, the simulation day is incremented and all the game variables are updated).

This work was previously published in Handbook of Research on Serious Games as Educational, Business and Research Tools, edited by Maria Manuela Cruz-Cunha, pp. 135-156, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 43
Serious Gaming Supporting Competence Development in Sustainable Manufacturing

Heiko Duin
BIBA – Bremer Institut für Produktion und Logistik GmbH, Germany

Gregor Cerinšek
Institute for Innovation and Development of University of Ljubljana, Slovenia

Manuel Fradinho
The Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology, Norway

Marco Taisch
Politecnico di Milano, Italy

ABSTRACT

Becoming a sustainable global manufacturing enterprise is a challenge for almost every manufacturing organization in the world because of its multidimensional nature. Sustainability combines environmental, economic, and social dimensions and is considered to be a complex and hard to learn subject needing a lot of experience and competences. Traditional ways to create such experience and develop competences like role playing and simulations tend to take a lot of time and are expensive. On the other hand, serious gaming has proven to support learners in acquiring new and complex knowledge and is ideally suited to support problem based learning by creating engaging experiences around a contextual problem where users must apply competences to solve these presented challenges. This chapter introduces a new learning environment which is build around a gaming engine supporting the development of competences in specific subject areas. Selected competences in sustainable global manufacturing lead to the definition of scenarios, which then can be executed by a game engine, thus creating experience within the user. A knowledge ecology space allows the user to interact and reflect on learning outcomes with other participants. The subject of sustainable global manufacturing is the application case presented in this chapter showing how specific competences in this area have been identified and how a game scenario has been developed. Finally, its implementation and evaluation is discussed.

DOI: 10.4018/978-1-4666-1945-6.ch043

INTRODUCTION

An early and broad definition for sustainable development has been provided by Gro Harlem Brundtland, former prime minister of Norway, during her presentation of the Report of the World Commission on Environment and Development to the UNEP's 14th Governing Council Session in 1987: *"Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs".* Overall sustainability includes the balancing of three dimensions: Environmental, economical and social sustainability.

In order to handle such a complex subject the main competence of involved managers is most probably the ability of multi-dimensional thinking, which is a competence hard to teach. Gaining such a competence relies strongly on experience. Traditional teaching (and learning) methods would focus on a one-to-one case study (or role play simulations) which is time consuming and, therefore, quite expensive in the end. The question is how such competences can be build in a learner with less time and less efforts concerning teaching staff.

There is a growing interest in games for purposes beyond entertainment. There is a common agreement that Serious Games are potential facilitators of improved learning whilst at the same time entertaining and engaging users. Examples have been developed in a wide range of fields including management science, economics, intercultural communication, psychology, military strategy, sociology, political science, and interpersonal skill development (Raybourn & Bos 2005). The hypothesized benefits of digital game-based learning include increased motivation of learners, and improved transfer of learning to the context in which the learning is applied. There is growing recognition of game based learning as a useful tool to support learning within schools and universities (Eck 2006).

However, a key challenge of building a Serious Game as an effective learning tool is the need to tailor the game design to address the competence domain. Although important, a serious game alone does not suffice to ensure learning and it is necessary to have a set of methodologies and tools to provide a learning environment that enable the tailoring of personalized learning plans tailored to the needs of the learner and supporting the individual (and the teacher) in assessing the progress made. Finally, understanding of a subject is often gained through reflection and exchange with other learners.

This chapter follows several objectives. First, the complex subject of Sustainable Global Manufacturing (SGM) is introduced and motivated. A detailed literature research provides insights in the competences one need to have to handle issues related to SGM. Based on those competences a game scenario which is part of a larger learning platform is introduced supporting learners to create such competences. Finally, the implementation and evaluation of that game scenario in the context of the learning environment is discussed.

WHY THERE IS A NEED FOR SERIOUS GAMING IN SGM

Manufacturing industries account for a significant part of the world's consumption of resources and generation of waste. Worldwide, the energy consumption of manufacturing industries grew by 61% from 1971 to 2004 and accounts for nearly a third of today's global energy use. Likewise, they are responsible for 36% of global carbon dioxide emissions (IEA 2007).

Manufacturing industries nevertheless have the potential to become a driving force for the creation of a sustainable society. They can design and implement integrated sustainable practices and develop products and services that contribute to better environmental performance. This requires a shift in the perception and understanding of industrial production and the adoption of a more holistic approach to conducting business (Maxwell, Sheate et al. 2006).

Sustainable Manufacturing is part of a larger concept of sustainable development, which emerged in the early 1980's in response to increased awareness and concern over the environmental and social impact of economic growth and global expansion of business and trade.

The concept of sustainable development implies firstly the integration of environmental issues with the imperatives of economic development in order to meet the immediate needs of populations today without undermining the aspirations of future generations. However, the definition of the term "sustainable development" has been expanded to include the ideas of fairness and interdependence, not only between generations, but between the countries and peoples of the Earth. Social, cultural, economic and natural environments, whose harmonious development is essential to the welfare of humanity and of nature, are also included in the concept (WCSD 2010).

The three main ingredients of the sustainability are environment, economy and society.

- The **Environmental Sustainability** is the ability to sustain quality of the physical environment. It concerns issues as exploitation of renewable and non-renewable natural resources, preservation of quality of water, air and soil, energy efficiency, waste management, balance of eco-systems and preservation of biodiversity.
- The **Economic Sustainability** concerns sustainable human development and growth, leading to improved quality of life, employment, social and economic equity. Moreover, it requires improvements and better exploitation of technologies, knowledge and innovations as source of cost optimization and resource efficiency.
- The **Social Sustainability** is the relationship between development and current social norms. It implies distribution of wealth and services within and between generations. An activity is socially sustainable if it conforms to social norms or does not stretch them beyond the community's tolerance for change. Social norms are based on religion, tradition, and custom; they are rooted in values attached to human health and well-being.

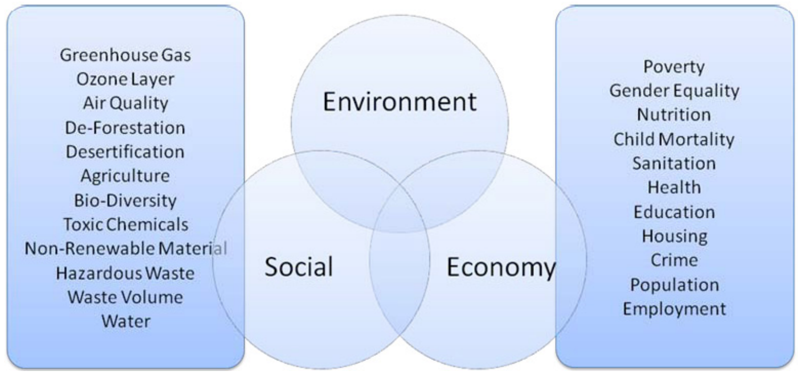
Sustainable development is used more often as an umbrella term, including many multidisciplinary areas of human activity. Some of them are: Urban regeneration, Community development, Energy, Transport, Ecology, Pollution, Climate change, Peak oil, Resource efficiency, Carbon foot printing, Workforce development, Organizational learning, Corporate responsibility, Human rights, Social justice. Sustainable development is emerging as basic principle for development in several industries as: sustainable building, sustainable transportation, sustainable manufacturing, sustainable utilities and services (WCSD 2010).

Sustainable Manufacturing is a business practice of the industrial sector, which expands all the company's processes and decisions into the broader social and natural environments in which the company operates in and affects with its actions. This is executed with the explicit objective of reducing or eliminating any negative impact, while pursuing the desired level of technological and economic performance (Leahu-Aluas 2010).

The US Department of Commerce defines Sustainable Manufacturing as the creation of manufactured products that use processes that minimize negative environmental impacts, conserve energy and natural resources, are safe for employees, communities, and consumers and are economically sound (Leahu-Aluas 2010).

Sustainable manufacturing is more comprehensive and systemic than e.g. green, eco-manufacturing, eco-machining, clean production by dealing with all three components of sustainability: environment, economy and society. While it

Figure 1. The three dimensions of sustainability and United Nations indicators



includes all the environmental concerns, such as pollution, material toxicity, GHG emissions, it is not limited to those concerns, nor is it a component of an environmental management system. Sustainable manufacturing uses both technological and non-technological solutions, from selection of materials and production processes to organizational mission, structure and performance reporting. It shifts the focus from "end-of-pipeline" solutions, disposal of waste, clean-up, recovery, a liability approach, to the very beginning, at product or process design stage, an opportunity approach (Leahu-Aluas 2010).

It resolves some of the limitations and controversies of an exclusively green/environmental view by aiming for a balance in satisfying all three components of sustainability, by aiming for a "triple bottom line" and a "triple top line". Unlike some green/environmental models and concepts, sustainable manufacturing builds upon and improves an advanced and technologically sophisticated industrial base. It recognizes the weakness of economic growth which disregards the limits of natural resources and the demands of the society for sharing the benefits of successful enterprise. Another important distinction is the inclusion of time, as a component of sustainable manufacturing, from extending the viability and longevity of the business to assessing and owning the entire life-cycle of a product or service (Leahu-Aluas 2010).

Sustainable manufacturing uses a more holistic approach that considers all life-cycle stages, from pre-manufacturing, manufacturing and use through post-use. This holistic view of sustainable manufacturing takes the aim of adopting sustainable principles across the whole manufacturing life-cycle, and identifying all inputs and outputs for continuous improvement action. These stages are spread across the entire supply chain with different partners managing activities at each of these stages. Thus, many players in the manufacturing process must adopt sustainable principles to ensure that higher production standards are met (WCSD 2010).

From historical perspective, different objectives for education that address environmental and sustainability challenges have been documented for the first time at the Intergovernmental Conference on Environmental Education organized by UNESCO in Tbilisi in 1977, also known as the "Tbilisi declaration" (ICEE 1977). Furthermore, Education and competence development for Sustainable Manufacturing and Sustainable Development is perceived as a matter of global importance and has been one of the top priorities in national policy documents and on the global agenda since the UN's Earth Summit in Rio in 1992 (UN 1992) and again brought to the forefront of the international attention at the UN World Summit on Sustainable Development in Johannesburg in 2002 (UN 2002). Both reports argue that education is critical for promoting sustainable development issues which should be integrated in all disciplines, should employ formal and informal methods and effective means of communication (UN 1992; 2002).

According to several national strategies and reports (e.g. Forum for the future 2004; NCS 2007; ASCE 2009; NAE 2004; Higher Education Academy 2008) the engineering profession can and should play an important role in securing a sustainable future. Engineering educators around the world are witnessing a significant shift in societal expectations of the engineering profession, to help address immediate and longer-term sustainable manufacturing challenges. Fokkema et al. (2005]) stated that "*each engineer has a responsibility to society and should have an awareness of possible ethical, social, environmental, aesthetic, and economic implications of their work and to act accordingly.*" As there is still a common global concern about the lack of sustainability content in engineering curriculum (e.g. Desha et al., 2009; Holmberg et al., Azapagic et al., 2005) trends of introducing competence-based sustainability-oriented education and training are becoming more and more apparent (e.g. NCS, 2007; Lourdel et al., 2005; Fenner, et al., 2005; Fokkema et al., 2005; Barth et al., 2007; Desha et al., 2009; Anderberg et al., 2009).

There is a broad range of research described in the literature dealing with competences in the field of sustainable manufacturing and sustainable development. De Haan (2010) proposed a model of "Gestaltungskompetenz" that centers on defining key competencies and related sub-competencies in the field of sustainable development in the education sector. Gestaltungskompetenz can be split into 12 sub-competencies, namely the ability to

1. gather knowledge in a spirit of openness to the world, integrating new perspectives;
2. think and act in a forward-looking manner;
3. acquire knowledge and acting in an interdisciplinary manner;
4. deal with incomplete and overly complex information;
5. co-operate in decision-making process;
6. cope with individual dilemmatic situation of decision-making;
7. participate in collective decision-making process;
8. motivate oneself as well as others to become active;
9. reflect upon one's own principles and those of others;
10. refer to the idea of equity in decision-making and planning actions;
11. plan and act autonomously; and
12. show empathy for and solidarity with the disadvantaged.

The Australian Manufacturing Skills Council imported the Guideline Competency Standards for Sustainability and they are designed to guide teaching and learning of respected competences in the field of sustainable manufacturing in formal vocational training and education settings as well as on the job (NCS 2007). The required knowledge, skills and relevant performance criteria required for the upper management level are structured under the competitive manufacturing unit "Develop workplace policy and procedures for sustainability". This unit is furthermore divided into four main competence elements:

1. Develop workplace sustainability policy;
2. Communicate the policy;
3. Implement the policy;
4. Review the policy implementation.

The Institution of Civil Engineers (ICE) established a Task Group to drive the implementation of sustainability principles into education, training and professional development (Fenner et al., 2005). It proposed the following broad skill requirements in the domain of sustainable development:

1. The ability to work with complex/ill-defined problems;
2. The development of team work and communication skills, and
3. The ability to evaluate the merits and demerits of options.

The Royal Academy of Engineering (RAE) identified a set of principles of Engineering for Sustainable Development (Fisk and McQuaid, 2004). Engineers are required to have the ability to:

1. operate and act responsibly, taking account of the need to progress environmental, social and economic issues simultaneously;
2. use imagination, creativity and innovation to produce products and services which maintain and enhance the quality of the environment and community and meet financial obligations, and,
3. understand and encourage stakeholder involvement.

Engineering Department at Cambridge University proposed essential competence ingredients for well informed engineers to tackle the sustainability issues (Fenner et al., 2005). They can be summarized as:

1. the need to engage in problem definition through careful dialogue with all stakeholder groups and a proper recognition of context;
2. an understanding of mechanisms for initiating and managing change in organizations so future engineers are equipped to play a leadership role;
3. an acknowledgment that technical innovation and business skills also must be understood, nurtured and combined as precursors to the successful implementation of sustainable solutions.

Forum for the future (2004) established a Higher Education Partnership for Sustainability that identified a sustainability literate person, who would be expected to:

1. Understand the need for change to a sustainable way of doing things, individually and collectively;
2. Have sufficient knowledge and skills to decide and act in a way that favors sustainable development;
3. Be able to recognize and reward other people's decisions and actions that favor sustainable development.

These explications show that Sustainable Global Manufacturing is a complex and hard to learn subject needing different competences and a lot of experience, while at the same time it is important for the industry to continue into this direction. There is a strong need that students practice and learn about sustainable manufacturing areas which may be currently underspecified in conventional courses (e.g. holistic and systems thinking, decision making for sustainability, interdisciplinary cooperation, communication skills etc.); new to the manufacturing industry (e.g. life cycle analysis, Eco-Management and audit scheme etc.) or ones that pose new challenges that have not been foreseen earlier. Consumers furthermore expect to be offered products which have low impact on environment and being socially acceptable (e.g. exclusion of child work).

CHALLENGES IN SUSTAINABLE MANUFACTURING

Challenges for Globally Acting Manufacturing Enterprises

Manufacturing enterprises which want to introduce Sustainable Global Manufacturing do face several challenges. First of all, the demand of customers for green products is still growing. There are several forces driving this trend: energy process are volatile, governments are interested in and try to regulate environmental issues, in public concerns about product and food safety do exist, and finally the environment isn't getting cleaner just by its own. A recent study shows that 34% of European adults are systematically looking for and often purchase green products and 24% of European adults would accept a higher price for green products in 2008 (Manget et al, 2009).

As outlined above, Sustainable Global Manufacturing could only be reached by balancing the three dimensions of environmental, economic and social sustainability. Therefore, the main challenge when implementing sustainability is to deal with this multi-dimensional issue in a holistic way. Optimization of specific factors may result in negative impacts in other areas (e.g. to make a product more green may result in higher energy consumption of manufacturing processes).

Complex products are assembled form many parts delivered by lots of different suppliers. To assure the green quality of a product there is a need to manage all the delivery parts to be green as well. In some sectors, suppliers are organized in tiers, where the first tier suppliers are cooperating quite closely with the manufacture while second and third tier suppliers are handled by the first tier. The challenge is to manage all suppliers down this hierarchy to support sustainability in order to ensure sustainability on the top level.

Each product is following a life cycle running through three phases starting with BOL (Beginning of Life) via MOL (Middle of Life) to the EOL (End of Life). The BOL phase deals with product development and manufacturing, MOL is about product usage and maintenance, and finally the EOL phase is for disassembly, refurbishing, and recycling. Note, that a sustainable view on manufacturing includes the absence of waste, there is nothing which needs to be disposed or dumped throughout the whole life cycle because everything acts as input for other manufacturing processes (McDonough and Braungart, 2002). Here, the challenge is to make a product sustainable throughout its whole life cycle.

If an enterprise decides to offer green products and to be sustainable it is necessary to act sustainable in all matters. It wouldn't be a good idea to offer green products while the management staff is flying round the world using cheap airlines and thus contributing to unnecessarily to air pollution. Sustainability is a behavior which should be reflected by each single employee of that company otherwise the image of the enterprise concerning sustainability is beyond belief.

Finally, connected with sustainable manufacturing one can observe the emergence of many different standards. Examples for these include ISO 14001, GRI (Global Reporting Initiative), EMAS (Eco Management and Audit Scheme), LEED (Leadership in Energy and Environmental Design), PEFC (Program for Endorsement of Forest Certification Schemes), OHSAS (Occupational Health and Safety Assessment Series) and ENERGY STAR which is a product label for energy-saving equipment and buildings.

Summarizing this discussion, the main challenges for globally acting manufacturing enterprises are:

- Responding to the growing demand for green products
- Holistic or systemic view on sustainability, i.e. balancing the three dimensions
- Covering the whole supply chain by including suppliers, but not just first tier suppliers, also second and third tier suppliers
- Managing the process of selecting and introducing sustainability standards

Challenges for Developing Competences in SGM

Among the main identified disadvantages of the traditional engineering and manufacturing education is firstly the predominant orientation towards factual knowledge (Dolinsek et al, 2004;

O'Sullivan et al, 2009; Westera, 2001); and secondly the issue that curriculum is not reflecting the real needs of the business, manufacturing industry and society as a whole (O'Sullivan, 2009; IMS Zurich, 2007). It has been argued that traditional engineering and manufacturing education is still mainly focused on developing the production skills of engineering students (Lehmann et al. 2008). With the acknowledgement that the sustainable development of today's societies relies on more than this potential, it inherently becomes necessary for future employees and knowledge workers to contribute with other potentials and skills as well.

The research from Azapagic et al. (2005) demonstrate that students generally believe that knowledge on sustainable development is important for them, although they often have difficulties in making a direct link between the theory of sustainable development and practice. As employers demand for competent individuals that can effectively tackle with real-world complex problems it appears that the results from the accomplishment of learning tasks which support a traditional knowledge-based curriculum are no longer sufficient for these dynamic social and work conditions (Sampson and Fytros 2008; Westera, 2001). The concept of competence is highly associated with the ability to master such complex and ill-defined situations. As a result, the term has become attractive for both educators and employers because it is easily identified with valued capabilities, qualifications and expertise. It appears that the general understanding is that competences are the crucial explanatory factor in situations where the employee is theoretically equipped to handle a work situation but in practice fails. Therefore, competence-based education and training can be seen as a way forward in addressing this kind of "missing piece" required in successful work performance.

In order to remain competitive manufacturing enterprises need to recruit and train competent employees who will be able to operate in future sustainability-oriented environments and who will provide the knowledge and expertise to master real-life complex situations. As sustainable manufacturing is and will become a major player in the landscape of developed economies, these skills and methods of their acquisition and development will need to change. Problem based technology enhanced learning can add to those skills by providing and promoting the linkages to and the understanding of other items and dimensions of the sustainability potential through context-based, interdisciplinary learning. This can enable students to solve complex and situated problems. Therefore universities and other education and training institutions need to offer their students innovative learning environments that will facilitate their learning and competence acquisition (Wert 2004; IMS Zurich 2007; Sampson and Fytros 2008; Petersen and Heikura 2010). The Technology-enhanced competence-based learning (Sampson and Fytros 2008) is recently gaining much interest in educational settings as it can enable the delivery of successful courses that take full advantage of the benefits made possible by the technology.

The research conducted within the European FP7 Integrated project TARGET on competence identification and development, that was based on the classic full-scale 6 step version of a competence study (Spencer and Spencer 1993), critical incident technique (Flanagan 1954; McClelland 1973; Cerinsek et al. 2007; Cerinsek and Dolinsek 2009) and job analysis (Pearn and Kandola 1990),identified, that employees at upper management levels are involved in the development and implementation of strategic sustainability initiatives. Many sustainability problems require solutions/concepts that take a system and holistic view of the world, accounting for complexity and coupling inter-disciplinary understanding with a much broader concept of system performance. The attempt to understand complexity and relationships should be the starting point for the employees. They are furthermore required to demonstrate personal responsibility and autonomy in performing complex technical operations and to take full responsibility for personal and group outcomes. They are required to consult widely with different stakeholders in order to develop workplace sustainability strategy and procedures that are in alignment with the strategy. The generic competences with related sub-competences that have been identified in the field of sustainable manufacturing for employees at upper management levels are presented in a form of the following competence set:

1. Personal commitment to the sustainability
   - Corporate social responsibility awareness
   - Environmental care and green future orientation
   - Broader view and open mind; integrating new perspectives
   - Decision making for sustainability (every decision takes into account sustainability issues)
   - Understanding the nature, principles, drivers, barriers and impacts of sustainability
   - Taking initiative and actions that reflect commitment to sustainability
   - Ability to understand, cope with and manage change
   - Ability to raise general sustainability awareness
   - Communication skills
2. Systems / holistic thinking
   - The ability to take a multi-scale perspective,
   - Life-cycle thinking,
   - Ability to perceive environmental challenges and policies (regulations and standards) not as a barrier to growth but as a new opportunity
   - Ability to differentiate between short and long term options and strategies and their consequences,
   - Ability to critically value situations from sustainability perspective,
   - Analytic thinking and the ability to identify and consider particular smaller parts of the system (zoom in and breaking things apart)
   - Synthetic thinking and the ability to consider system as a whole, perceiving the big picture (zoom out and putting things together)
   - The ability to understand the interrelationships, cause-effect chains and trade-offs between different parts of a system
   - Conceptual thinking as seeking connections/patterns or inconsistencies/discrepancies that are not obvious to others
   - Ability to deal with complexity
   - Ability to make sense of a large amounts of complex and incomplete information and identifying key issues in complexity
   - A strong macro-ethic,
   - Ability to acquire knowledge and act in an interdisciplinary manner,
   - Ability to act in heterogeneous and multidisciplinary groups and utilize the diversity
   - Openness, creativity and flexibility in thought,
   - A strong business sense,
   - The ability to empathise with other people,
   - Communication skills (formal and informal),
   - An aptitude for lifelong learning
3. Development of a sustainability strategy and related policies
   - The ability to think and act in a forward-looking manner
   - Ability to define future requirements and interpret them into strategy

- Analytic and research skills (to research, identify and evaluate different sustainability policies, standards, legislation, regulations)
- Ability to understand and deal with different tools and methods that are addressing sustainability from various points of view
- Ability to manage cost implications of sustainability
- Communication and negotiation skills (to consult relevant stakeholders)
- Social networking skills
- Ability to introduce new business model
- Ability to plan cross discipline work
- Ability to search for relevant feedback
- Decision making
- Creativity and resourcefulness
- Ability to cope with change
- Readiness to compromise
- Ability to understand emerging market trends

4. Communication of a sustainability strategy in the organization and fostering sustainability culture
   - Ability to embed sustainability into company processes and organization structures
   - Communication and negotiation skills (communicate and effectively present the strategy and other relevant messages to critical stakeholders)
   - Consultation, delegation and empowerment (people who will be implementing the strategy; employee involvement)
   - Ability to motivate, lead awareness and drive sustainability idea forward
   - Capability to generate enthusiasm and encourage cooperation
   - Ability to develop common understanding
   - Social networking skills
   - Ability to co-operate in decision-making process
   - Readiness to compromise
5. Implementation of a sustainability strategy
   - Goal orientation
   - Ability to work through obstacles to complete assignments
   - Communication skills (effectively communicate procedures to help implement the strategy)
   - Delegation and assigning responsibilities
   - Decision making
   - Ability to effectively manage possible trade-offs (triple bottom line)
   - Ability to identify needs and take appropriate actions
   - Ability to manage cost implications of sustainability
   - Demonstrating managerial courage
   - Time management
   - Ability to engage in complex processes
   - Ability to manage multiple complex projects
   - Ability to implement cross discipline work
   - Ability to work under pressure
   - Ability to deal with differences and contradictions (not rushing to a single solution but handles tensions)
6. Formation and leadership of multi-disciplinary teams (multidisciplinary cooperation)
   - Social networking skills (building multi-sectoral and multi-stakeholder networks)
   - Communication, verbal and negotiation skills (ability to listen to others)
   - Flexibility
   - Ability to build positive relationships with others
   - Ability to motivate

    - Ability to drive sustainability idea forward
    - Capability to generate enthusiasm and encourage cooperation
    - Ability to express complex issues when dealing with people coming from different backgrounds (in multidisciplinary teams)
    - Ability to utilize diversity in teams
    - Ability to take over different perspectives
    - Tolerance and acceptance with regard to other disciplines
    - Demonstrating managerial courage, leadership
    - Problem solving and conflict resolution
    - Ability to supervise the work of the team
    - Readiness to compromise
7. Review, evaluation and reporting of strategy implementation
    - Analytic skills
    - Ability to interpret the relevant data
    - Ability to provide feed back to key stakeholders
    - Ability to investigate success
    - Deliver rewards and recognition (e.g. other people's decisions and actions that favour sustainability)
    - Ability to drive results across multiple simultaneous projects

The main target groups addressed by the serious game described in the following section are young management professionals and engineering students involved in manufacturing relevant courses (e.g. MBA students with an experience level of two years). The serious game is applied in a blended learning concept: First, students get a series of lectures on sustainability followed with gaming sessions to deepen acquainted knowledge.

Serious Gaming seems to be the most promising approach to generate competences as mentioned above due to the complex and inter-linked competences. In the mid 90s of the 20th century, unaware of the term Serious Gaming, several research institutes started to develop games to be applied in the domain of manufacturing (Hauge et al. 2010). Most of these games have been developed either at one single institution or as a part of multi-national research projects. Furthermore, Hunecker (2008) presents a game that aims to help the player understand complex production systems. The game uses a resource-based process model in which a process transforms resources from one form to another. The payer's goal is to use the resources in a way that maximizes profit. Another serious game within the area of production systems is presented in by Duin et al. (2007). In this game the player manages a production system and competes on a global market. The game aims to teach young management staff strategic decision making in global production.

Serious games can be very effective for teaching sustainability to students as they not only convey hard/technical skills but also mediate soft skills like collaboration, creativity and communication (Scholz-Reiter et al. 2002; Hauge et al. 2010), thus tackling wicked problems that are characterized by a high degree of complexity, uncertainty and conflict and are not usually addressed by other learning methods. According to Digiesi et al. (2010) experiential and problem-based learning performed through a competitive serious game helps practitioners, engineers and master students in being aware of the overall problem complexity of sustainability (e.g. addressing the triple bottom line of sustainability).

A SERIOUS GAME MEDIATING COMPETENCES IN SGM

Requirements for the Serious Game

The methodology for collecting and defining user requirements for the Sustainable Manufacturing Scenario is based on two premises:

1. The requirements engineering is incremental and iterative. Therefore, each report just provides a snapshot on the results of requirements engineering for a specific point in time.
2. The methodology to elicit, document and manage user requirements which has been used in earlier phases of the TARGET project is also applied in the Sustainable Manufacturing area. Small modifications concerning the templates have been done. The methodology is based on the Volere template (Robertson & Robertson, 1999).

The requirements presented in the following chapter arise from two main sources:

1. Interviews with subject matter experts have been performed. Besides the collecting of information about necessary competences to implement SGM those experts were also asked to provide potential requirements for a system supporting the learning of such competences. These interviews were performed by various consortium members of the TARGET project.
2. The SGM game scenario is continuously developed and refined. A critical reflection and analysis of that scenario also provided valuable input for the requirements analysis.

The template shown in Table 1 has been used for elicitation and documentation of user requirements within the interviews. The same table has also been applied while analysing the GSM game scenario.

In total 45 requirements have been identified from which are 15 functional requirements, 22 usability requirements, three security requirements and three organizational requirements. Two requirements are related to costs. As for the date this chapter is provided not all of the scheduled interviews have been completed it is expected that further requirements will emerge in future. Table 2 provides some examples of identifies user requirements.

Methodology

Unlike entertainment games, serious game development relies on a close collaboration between game designers and developers and domain experts, as game designers and developers are usually not familiar with the pedagogical content, whilst domain experts rarely have the skills to design and build games (Eck 2006). Anecdotal

Table 1. User requirement template

Requirement No.	
Descriptive Name*	
Requirements Type	Functional Requirement Usability Requirement Security Requirement Organisational Requirement Cultural / Political Requirement Legal Requirement Costs
Version	
Author*	
Rationale*	
Description*	
State	
Feasibility	
Priority	
Comments	
* All fields marked with a * are mandatory.	

Table 2. Examples of user requirements

No.	Name	Comments
SGM-FU-02	Manufacturing Simulation	None
SGM-FU-03	Market Behavior in Simulation	None
SGM-FU-04	Sustainability Dimensions in Simulation	None
SGM-FU-05	Environmental Indicators in Simulation	None
SGM-FU-08	Resembling a Real Office	None
SGM-FU-09	Permanent Access to Data	Concerning simulated data.
SGM-FU-10	Staff Meetings (Virtual)	None
SGM-US-01	Positive User Experience; Responsive Environment; providing hints when player get stuck	None
SGM-US-02	Accurate Virtual Reality	This raises security issues when the environment of the customer is accurately modeled.
SGM-US-04	Real Life Examples (which happened already at the customer)	None.
SGM-US-05	Provision of SGM Knowledge	Provision of learning materials.
SGM-SE-01	Anonymity	Could be conflicting with SGM-OR-01, SGM-OR-03
SGM-SE-02	Secure Access to the TARGET System	None
SGM-OR-01	Integration of Colleagues / Partners (Multiplayer)	None

evidence suggests that gathering and using information via traditional mechanisms such as focus groups and interviews for serious game design can be problematic. For example, in the case of the Virtual Leader serious game, plans for close collaboration with domain experts were eventually dropped when it became clear that the insights yielded would be difficult to integrate into the game. In the end, the game design relied on a crafted approach confined to the ideas of the developers (Aldrich 2009).

The adopted methodology (Seager et al 2010) to devise the serious game in the domain of sustainable manufacturing is based on a structured approach to game design for serious games by extending the Applied Cognitive Task Analysis (ACTA) method, resulting in Table 3.

The Game Scenario

The game scenario reflects the phases an enterprise has to run through when dealing with sustainability issues. The first phase – the initialization phase – is about identification of problems and issues concerning sustainability. The enterprise checks whether it is already compliant with sustainability standards or whether there is potential for improvements. An outline of new sustainable processes within the enterprise is developed. During the second phase different options concerning process standards, product standards, and other company wide standards are evaluated. Examples for standards are: ISO (International Standardization Organization), EMAS (Eco-Management and Audit Scheme), or GRI (Global Reporting Initiative). During the third phase a decision has to be taken about what standards to apply or introduce in the organization. The fourth phase is the realization or implementation of the decision follow by the last phase which concerns reiteration, i.e. continuous improvement on the selected standard and restarting the process with other standards.

Enterprises dealing with sustainability issues often apply a standard called Life Cycle Analy-

Table 3. Steps of the methodology

Step	Description
1	Determine who the target learner is
2	Determine the overarching competence to be captured by the game
3	Breakdown the chosen competence into critical incidents
4	For each critical incident: Identify main task and the main decisions that a learner must do Explore various alternatives Identify what is difficult and challenging in those decisions Identify typical mistakes made by a novice, along with cues/strategies to deal with mistakes If critical incident involves more than one character, identify how they can be used to shape the situated context towards achieving the desired learning outcomes
5	Define the task model
6	Identify the belief, desire and intentions for the agents representing the different characters

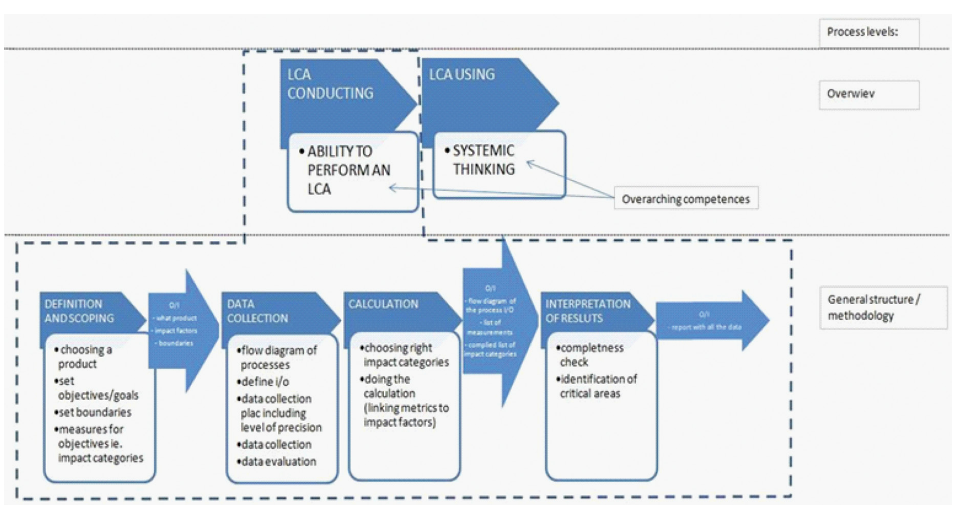
sis (LCA) to identify environmental impacts of their organization, products or processes (Curran, 2006). Generally, the application of LCA includes two major phases (see Figure 2):

1. Conduction the LCA, i.e. defining the scope of the analysis, collecting necessary data, to calculate the environmental impacts and to evaluate and interpret the results.
2. Using the results of the LCA to decide on enhancement measures to reduce the environmental impact.

This LCA process serves as our first case in the context of the SGM game scenario. The overarching competences addressed by this case are the ability to perform an LCA and systems thinking.

Within the game scenario the player takes over the role of a Sustainability Manager who was recently hired by the CEO (Chief Executive Manager) of a production company. When starting the game scenario the player finds himself in a meeting with the CEO and other important managers (production manager, logistics manager, human resources director, etc.). The CEO introduces the

Figure 2. The life cycle assessment (LCA) case



plan that a LCA should be conducted concerning a specific product and advises the player to do so. He also urges the other managers to support him. The CEO and the other managers are non-player characters who are driven by a sort of artificial intelligence of the game engine. After the meeting finishes the player – who is already familiar with the process of conducting a LCA – starts to execute the four phases.

In the first phase (scoping of the LCA) the player has to define the objectives and goals of that LCA. He can do so by just deciding on the information he got from the CEO or by interviewing other managers. On whatever he decides it will have some impact on costs, time and the final quality of the LCA. The next step for the player is to decide on the boundary, i.e. focusing on a specific department, or focusing on the production, or focusing on the whole supply chain or analyzing the whole life cycle of the selected product. The player is supported by a virtual LCA Tool where he enters the decisions he has made.

During the second phase (data collection) the player needs to draw up the process(es) necessary to produce the selected product. A process step is considered as a black box which has some inputs (materials, parts, energy, etc.) and outputs (parts, products, waste water, pollution, etc.). All process steps are put together to a flow diagram which determines what data needs to be collected. Again, the player has virtually any freedom to do this job. He can e.g. ask the production manager about the processes, he can talk to the shift managers or he is visiting the shop floor by himself to see how production is working. After identifying what data is needed he is going to collect that data by asking colleagues to provide it. The results are entered in the virtual LCA tool.

The third phase is about the calculation of environmental impacts. To do so, the right impact categories need to be selected. The final calculation is done by the virtual LCA tool which also reports whether all necessary data has been entered or not.

Finally, the fourth phase is checking the completeness and consistency of all collected data and the evaluating the results in terms of impacts per category. The final result is a report to be created with the virtual LCA tool and delivered to the CEO. Connected with this report is the costs and time which was needed to generate that report and its quality (meaning the right focus, correctness of processes and collected data).

For inexperienced users of the LCA method there are several opportunities to make mistakes while executing the conduction of a LCA. The following Table 4 provides an overview on the challenges a player is confronted with when conducting a LCA.

Typical mistakes a novice in LCA can make include:

- Setting wrong goals and objectives, selecting too little or too many parameters
- Ending with an unbalance between overall quality, time and costs in conducting the LCA project
- Choosing planned vs. real process execution facts
- Too broad or too focused boundaries for the LCA
- Selection of wrong or not relevant impact categories for the LCA
- Lack of holistic view of interviewed persons: usually people's knowledge does not cover overall end-to-end processes
- Lack of holistic and systematic view of player
- Incomplete or wrong definition of the product production processes
- Incorrect definition of the unit of measurements
- Incorrect use of measurement units
- Biased measurements
- The player does not understand each of impact category
- The player has focused too much or too less in details of each impact category

Table 4. Challenges for the player while conducting a LCA

Phases	Challenges and Opportunities for Mistakes
Phase 1 Definition and Scoping	The selection of objectives and goals (parameters) has direct impact on the result of LCA. A wrong choice of parameters can have negative impacts on the overall quality, time and costs of the LCA project. The player has the challenge of getting information from the right source of information. The player should choose from whom he wants to get the information. The player needs to define the scope of LCA in order to outline how to proceed, make efficient use of time and resources, the final results obtainment and better identification of boundaries. The system boundary determines which unit/department and which processes are included in the LCA and must reflect the goals of the project.
Phase 2 Data Collection	The determination of inputs and outputs in each process will help the player to better understand what data is required to collect. The player needs to define the measurement units correctly. Any incorrect or biased data will lead to wrong results and at the end it would negatively impact on quality, time and cost indicators of project. The determination and selection of the right impact categories for conducting the LCA could influence the overall time, quality and cost of LCA project.
Phase 3 Calculation	The whole LCA process is based on data and the player's ability to capture complete and accurate data. The players need to have the ability to dynamically re-scope the project in case of finding any insufficiency in the captured data. In the real world there are many possibilities to conduct the LCA process. Finding the most efficient order is a challenge for player.
Phase 4 Interpretation of Results	The interpretation of results is difficult in regard to the appropriateness and correctness of data in relation to the parameters and the impact categories of the LCA.

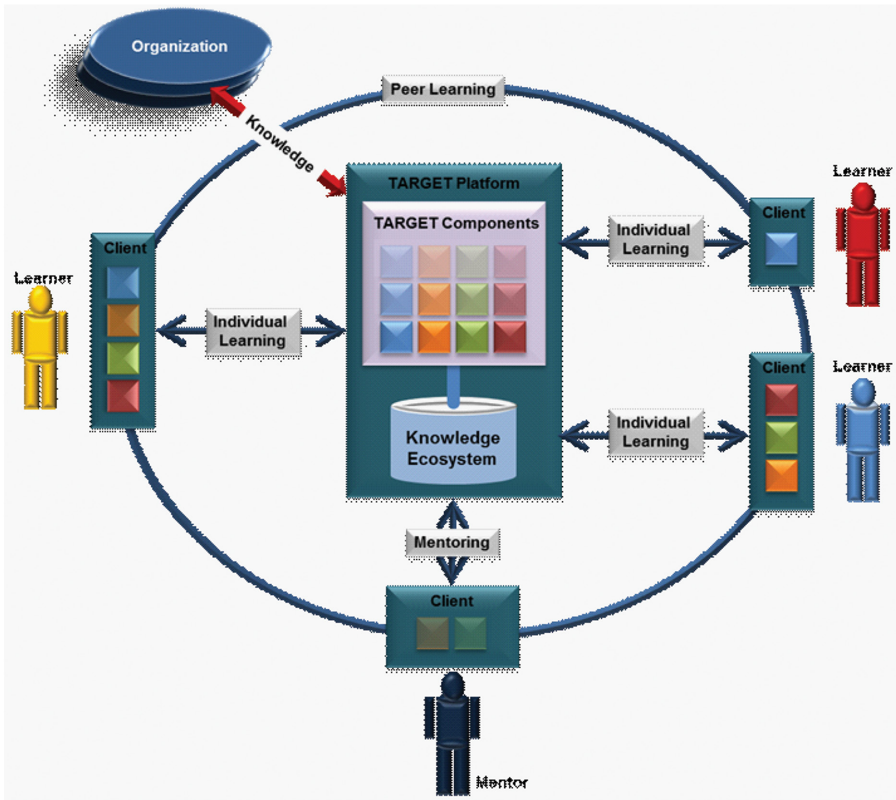
- The player has no idea of availability of data
- The player has problems with interacting with involved people and getting information from them
- The player just conducts desk work instead of going and visiting the shop floor level
- The number of conducted interviews is not adequate
- The player changes his mind too frequently and too often re-scopes the project
- The player does not re-scope the project when needed
- The player cannot decide on the quality of captured data
- The player couldn't find the most efficient order in executing the necessary steps of a LCA
- The player has a too narrow view and cannot see the different possibilities in conducting each phase

Implementation and Discussion of the Game Scenario

The Transformative, Adaptive, Responsive and enGaging EnvironmenT (TARGET) project aims to develop a novel Technology Enhanced Learning (TEL) platform that provides learners with a responsive environment that addresses personalized rapid competence development and sharing of experiences. The approach taken reflects upon James Baldwin's paradox (Baldwin, 1911) "*in order to be social you have to be individual and in order to be individual you have to be social*", thereby treating learning as a non-linear process moving away from a dichotomy of individual and social aspects of learning, whilst emphasizing dialogue and transformation (e.g. Threshold Concepts Framework (Flanagan et al, 2010)). An overview of the TARGET platform and its main stakeholders is captured in the diagram of Figure 3.

The resulting platform is based on an open architecture that supports different learning domains, with the project delivering support for the

Figure 3. Overview of the TARGET platform and its main stakeholders



domains of global sustainable manufacturing with strong focus on soft skills. In TARGET, the learning process draws heavily from Problem Based Learning (PBL) (Barrows and Tamblyn, 1990) and Action Learning (AL) (Gabrielsson and Politis, 2010), resulting in the use of digital interactive stories that provide situated rich contexts where a learner is required to apply and develop competences to achieve successful outcomes.

The approach taken in TARGET was to avoid developing another complete TEL platform, which would require the resulting platform to compete with an already established competitive market of existing solutions catering the needs of the enterprises concerning the human resource management landscape of services and tools. Consequently, the TARGET approach was to leverage the Service Oriented Architecture (SoA) alongside the "mash-up" paradigm to define an open framework of components can be put together in different destination platforms. As an example, one may consider the ROLE1 platform or even a well-established solution like iGoogle as feasible destination platforms for deployment of the TARGET platform.

The instantiation of the TARGET Learning Process is illustrated in the diagram of Figure 4, which can be characterized in the following stages:

- **Competences**. The TARGET Learning Process begins with the learner deciding on what competences to develop. This is done in one of two ways, either goal-oriented or self-directed learning. In the case of goal-oriented learning, the learner defines their current competence profile and their desired learning outcome in the form of outlining their target competence pro-

Figure 4. Overview of the TARGET learning process

file. The result of profiling leads to the creation of a learning plan based on custom stories tailored to the particular needs of the learner. Each story captures a business context, which may also involve defined characters with particular roles. The process of creating the learning plan is governed and shaped by a learning strategy that is chosen by the learner. In the case of self-directed learning, the learner builds their learning plan from the experiences of others within the community and these are stored in the knowledge ecosystem.

- **Story**. The learner is provided a background to the Story, which gives insight into the context, including the various characters available and their role in the Story. Some of the characters are available to the learner to be played by them, but in many cases the characters are only manipulated by non-learner character.
- **Experience**. Whilst engaged with the Story, the system provides an environment where the learner engages with other characters (either controlled by another learner or a Non-Player Character) and the environment, enacting their decisions. These decisions will have an impact which will affect and change the situated context of the Story. By monitoring the actions of the learner and taking into account the desired learning outcomes, the TARGET platform makes changes to the Story if necessary. As examples, these changes may be modifying the personality of a Non-Player Character to be more confrontational or delaying tasks within a project.
- **Reflection**. The learner is presented with the assessment of their competence during the experience in the form of a timeline manner. The ability of looking back on their decisions by reviewing how the story unfolded whilst cross-referencing the assessment of their competence at each point in time, allows the learner to evaluate their performance leading to reflection.
- **Peer Learning**. The TARGET learning process supports the learner in externalizing the tacit knowledge acquired after their experience of a Story, thereby contributing

to the creation of knowledge assets that are uploaded to the Knowledge Ecosystem. Once uploaded, the learning community plays an important role in the process with the support of recognized mentors as facilitators and in discussion with other learners. The social aspects address the need of an ability to deal with flux and instability, and to thrive in situations of flux (Kelly 1994).

Each of the four phases of the TARGET learning process is supported by a set of well-defined services embodied into components that are event driven, thus loosely decoupled from one another with some sharing functional dependencies. This means that the TARGET platform need not be entirely deployed as an integrated solution, but only subsets of the supported functionality. However, one needs to ensure that those components sharing functional dependencies are deployed together otherwise they may be operational at run-time but not work as required.

The adopted architectural design enables the platform functionality to be extended from the baseline features developed within the scope of the TARGET project and to replace existing components with alternate ones, provided there is compliance to the interfaces and the communication mechanisms.

The TARGET platform is geared towards a relatively thin client, which manages the flow supporting the TARGET learning process and provides interfaces for the different components. So within the context of the project, the learner engages with the TARGET platform through a web-based client, which interacts with the different TARGET components residing on the backend server. The decision to adopt a web based solution minimises adoption barriers associated to the deployment of technological solutions in corporate environments. However, it is possible for an organization to consider the deployment that is not necessarily web-based.

FUTURE RESEARCH DIRECTIONS

Future research in the context of the competence development will involve further modeling of competences in a way that reflects their complexity by incorporating more of the contextual factors to competences and illustrate how these contextual factors and the competences themselves are related to each other. This will furthermore facilitate the design of appropriate personalized learning plan and to associate specific competences to appropriate game scenarios. Additional application of the competence model in the field of sustainable global manufacturing will allow the identification of operationalized and measurable performance indicators taking identified contextual factors into account. The identified performance indicators and observing their occurrence within the context of the serious game will allow provide the evidence about the presence of the specific competences and sub-competences and therefore allow effective assessment.

While the implementation of the game scenario of performing a Life Cycle Assessment within the TARGET platform continues, first insights and experiences are quite promising. Therefore, besides simple testing an evaluation on a larger scale needs to be performed to collect user feedback. The objective of the evaluation is twofold: First, the believability of the game scenario has to be proven by involving subject matter experts (i.e. managers who have already substantial experience in performing an LCA in a corporate context). Second, expectations on competence enhancement in learner target groups need to be evaluated in comparison with other methods. User feedback will be interesting especially for the evaluation of the personal progress in that field (i.e. how engaging is the game and does the cognitive load balancing really works).

The development process in the TARGET project is based on agile principles. It is expected that new requirements will emerge while the implementation of the system matures. Also, the development of other game scenarios, e.g. based on another standard like the GRI, will lead to new requirements coming from the stories to be developed.

CONCLUSION

This chapter introduced the topic of Sustainable Global Manufacturing and related competences as an application area for Serious Gaming. While traditional manufacturing problems have already been addressed by serious games, the application area of Sustainable Manufacturing is new in this field.

The intended platform for game execution is called TARGET and combines functionality for mentoring and individual as well as peer learning. The TARGET platform allows the management and development of competences through story generation, experience acquisition and peer learning through reflection. As this platform is still under development only early reflections from pilot users are available but not yet formally recorded. The tenor of feedback received so far was positive. However, a large scale evaluation of the TARGET Platform and the SGM Scenario still has to be performed.

ACKNOWLEDGMENT

This work has been partly funded by the European Commission through the Project TARGET under contract No IST-FP7-231717. The authors wish to acknowledge the Commission and all project partners for their support.

REFERENCES

Aldrich, C. (2009). *The complete guide to simulations and serious games: How the most valuable content will be created in the age beyond Gutenberg to Google*. Jossey Bass.

Anderberg, E., Norden, B., & Hansson, B. (2009). Global learning for sustainable development in higher education: recent trends and a critique. *International Journal of Sustainability in Higher Education*, *10*(4), 368–378. doi:10.1108/14676370910990710

ASCE. (2009). *Code of ethics*. Retrieved October 15, 2010, from http://www.asce.org/uploadedFiles/Ethics_-_New/Code%20of%20Ethics%20October%202010.pdf

Azapagic, A., Perdan, S., & Shallcross, D. (2005). How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum. *European Journal of Engineering Education*, *30*(1), 1–19. doi:10.1080/03043790512331313804

Baldwin, J. (12911). Social institutions: The school, the state, the church. In R, G. Badger (Ed.), *The individual and society or psychology and sociology,* (pp. 118-144).

Barrows, H., & Tamblyn, R. (1990). *Problem-based learning: An approach to medical education*. New York, NY: Springer Publishing Company.

Barth, M., Godemann, J., Rieckmann, M., & Stoltenberg, U. (2007). Developing key competencies for sustainable development in higher education. *International Journal of Sustainability in Higher Education*, *8*(4), 416–430. doi:10.1108/14676370710823582

Brundtland, G. H. (1987, 8 June). *Report of the World Commission on Environment and Development to the UNEP's 14th Governing Council Session.* Nairobi, Kenya.

Cerinsek, G., & Dolinsek, S. (2009). Identifying employees' innovation competency in organisations. *International Journal of Innovation and Learning*, *6*(2), 164–177. doi:10.1504/IJIL.2009.022811

Cerinsek, G., Dolinsek, S., & Oliveira, M. (2007). *Using critical incident technique with serious games as the next step in organizational competence development*. Conference Learning with Games, Sophia Antipolis, France.

Curran, M. A. (2006). *Life cycle assessment: Principles and practice*. Report EPA/600/R-06/060 of the Scientific Applications International Corporation (SAIC), May 2006.

De Haan, G. (2010). The development of ESD-related competencies in supportive institutional frameworks. *International Review of Education*, *56*(2-3), 315–328. doi:10.1007/s11159-010-9157-9

Desha, C. J., Hargroves, K., & Smith, M. H. (2009). Addressing the time lag dilemma in curriculum renewal towards engineering education for sustainable development. *International Journal of Sustainability in Higher Education*, *10*(2), 184–199. doi:10.1108/14676370910949356

Digiesi, S., Mossa, G., Mummolo, G., Pilolli, R., & Ranieri, L. (2010). Learning sustainability through competitive game: Experiences from the Vinyl Chloride Monomer (VCM) case study. In M. Taisch, J. Cassina, & R. Smeds (Eds.), *14th Workshop of the Special Interest Group (SIG) on Experimental Interactive Learning in Industrial Management of the IFIP Working Group* (pp. 62-70), Milano. Retrieved October 15, 2010, from http://sites.google.com/site/ifip57sigexilim/

Dolinšek, S., Starcic, A., & Kopac, J. (2004). Education for manufacturing – From discipline to the competency based approach. In Taisch, M., Filos, E., Garello, P., Lewis, K., & Montorio, M. (Eds.), *International IMS Forum 2004* (pp. 1354–1362). Como, Italy: Global Challenges in Manufacturing.

Duin, H., Oliveira, M., & Saffarpour, A. (2007). A simulation model for virtual manufacturing environments for serious games. In K. S. Pawar, K.-D. Thoben, & M. Pallot (Eds.), *13th International Conference on Concurrent Enterprising*, Sophia Antipolis, France.

Eck, R. V. (2006). Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE Review*, *41*(2), 16–30.

Fenner, R. A., Ainger, C. M., Cruickshank, H. J., & Guthrie, P. M. (2005). Embedding sustainable development at Cambridge University Engineering Department. *International Journal of Sustainability in Higher Education*, *6*(3), 229–241. doi:10.1108/14676370510607205

Fisk, D., & McQuaid, J. (2004). *Principles of engineering for sustainable development*. London, UK: Royal Academy of Engineering.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327–358. doi:10.1037/h0061470

Flanagan, M., Taylor, P., & Meyer, J. (2010). Compounded thresholds in electrical engineering. In Land, R., Meyer, J. H. F., & Baillie, C. (Eds.), *Threshold concepts and transformational learning*. Rotterdam, The Netherlands: Sense Publishers.

Fokkema, J., Jansen, L., & Mulder, K. (2005). Sustainability: Necessity for a prosperous society. *International Journal of Sustainability in Higher Education*, *6*(3), 219–228. doi:10.1108/14676370510607197

Forum for the Future. (2004). *Higher education partnership for sustainability*. Learning and Skills for Sustainable Development Report. Retrieved October 15, 2010, from http://www.forumforthefuture.org/files/learningandskills.pdf

Gabrielsson, J., Tell, J., & Politis, D. (2010). *Business simulation exercises in small business management education: using principles and ideas from action learning*, (pp. 3-16). Routledge, Taylor & Francis Group.

Hauge, J. B., Riedel, J., Fradinho, M., & Westra, W. (2010). Addressing research fragmentation in serious gaming for manufacturing. In M. Taisch, J. Cassina, & R. Smeds (Eds.), *14th Workshop of the Special Interest Group (SIG) on Experimental Interactive Learning in Industrial Management of the IFIP Working Group* (pp. 62-70). Milano. Retrieved October 15, 2010, from http://sites.google.com/site/ifip57sigexilim/

Higher Education Academy. (2008). *Education for sustainable development in engineering*. Report of a Delphi Consultation. Engineering Subject Centre of the University of Manchester, UK. Retrieved October 15, 2010, from http://www.engsc.ac.uk/downloads/scholarart/delphi-consultation.pdf

Holmberg, J., Svanstrom, M., Peet, D.-J., Mulder, K., Ferrrer-Balas, D., & Segalas, J. (2008). Embedding sustainability in higher education through interaction with lecturers: Case studies from three European technical universities. *European Journal of Engineering Education*, *33*(3), 271–282. doi:10.1080/03043790802088491

Hunecker, F. (2008). A generic process simulation-model for serious games in manufacturing. *Proceedings of SAGSET 2008 – 38th Annual Conference: Teaching and Learning through Gaming and Simulation*. University of Nottingham, UK.

ICEE. (1977). *Intergovernmental Conference on Environmental Education in Tbilisi*. UNESCO Report. Retrieved October 15, 2010, from http://unesdoc.unesco.org/images/0003/000327/032763eo.pdf

IEA. (2007). *Tracking industrial energy efficiency and CO2 emissions*. Paris, France: OECD/IEA.

IMS Euro Swiss MTP Workshop. (2007). Swiss Federal University of Technology (ETH), Zurich, Switzerland.Retrieved October 15, 2010, from http://cordis.europa.eu/ims/past001_en.html

Kelly, K. (1994). *Out of control: The new biology of machines, social systems and the economic world*. New York, NY: Basic Books.

Leahu-Aluas, S. (2010). *Sustainable manufacturing – An overview for manufacturing engineers.* Resource document. Sustainable Manufacturing Consulting. Retrieved October 15, 2010, from http://sustainablemanufacturing.biz/media//DIR_17101/9543103a2878dfb8ffff83b8ffffd524.pdf

Lehmann, M., Christensen, P., Du, X., & Thrane, M. (2008). Problem-oriented and project-based learning (POPBL) as an innovative learning strategy for sustainable development in engineering education. *European Journal of Engineering Education*, *33*(3), 283–295. doi:10.1080/03043790802088566

Lourdel, N., Gondran, N., Laforest, V., & Brodhag, C. (2005). Introduction of sustainable development in engineers' curricula – Problematic and evaluation methods. *International Journal of Sustainability in Higher Education*, *6*(3), 254–264. doi:10.1108/14676370510607223

Manget, J., Roche, C., & Münnich, F. (2009). *Capturing the green advantage for consumer companies*. The Boston Consulting Group.

Maxwell, D., Sheate, W., & van der Volst, R. (2006). Functional and systems aspects of the sustainable product and service development approach for industry. *Journal of Cleaner Production*, *14*(17), 1466–1479. doi:10.1016/j.jclepro.2006.01.028

McClleland, D. C. (1973). Testing for competence rather than for "intelligence". *The American Psychologist*, *28*(1), 1–14. doi:10.1037/h0034092

McDonough, W., & Braungart, M. (2002). *Cradle to cradle: Remaking the way we make things*. New York, NY: North Point Press.

NAE. (2004). *The engineer of 2020: Visions of engineering in the new century*. Washington, DC: The National Academies Press.

NCS. (2007). *Standards for sustainability*. Resource manual, National Centre for Sustainability at Swinburne University of Technology. Retrieved October 15, 2010, from https://www.skillsonline.net.au/clearhse/Preview.do?no=216&type=O

O'Sullivan, D., Rolstadas, A., & Filos, E. (2009). Global education in manufacturing strategy. *Journal of Intelligent Manufacturing*, *22*. doi:doi:10.1007/s10845-009-0326-2

Pearn, M., & Kandola, R. (1990). *Job analysis: A practical guide for managers*. London, UK: Institute for Personnel Management.

Petersen, S., & Heikura, T. (2010). *Modelling project management and innovation competences for technology enhanced learning*. EChallenges 2010, Warsaw, Poland.

Raybourn, E., & Bos, N. (2005). *CHI '05 Extended Abstracts on Human Factors in Computing Systems, Portland, OR*. New York, NY: ACM.

Robertson, S., & Robertson, J. (1999). *Mastering the requirements process*. Addison-Wesley.

Sampson, D. G., & Fytros, D. (2008). Competence models in technology-enhanced competence-based learning. In H. H. Adelsberger, Kinshuk, J. M. Pawlowski & D. Sampson (Eds.), *International handbook on Information Technologies for education and training*, 2nd edition (pp. 1-25). New York, NY: Springer.

Scholz-Reiter, B., Gavirey, S., Echelmeyer, W., Hamann, T., & Doberenz, R. (2002). Developing a virtual tutorial system for online simulation games. *Proceedings of the 30th SEFI Annual Conference*, Firenze, Italy.

Seager, W., Ruskov, M., Sasse, A., & Fradinho, M. (2010). *Systematic approach to game design in project management and innovation.* 13th International Workshop of the Special Interest Group on Experimental Interactive Learning, Milan, June, 2010.

Spencer, L., & Spencer, S. (1993). *Competence at work: Models for superior performance*. New York, NY: John Wiley & Sons.

Taisch, J., Cassina, R., & Smeds, R. (2010). *Experimental learning on sustainable management, economics and industrial engineering*. 14th Workshop of the Special Interest Group (SIG) on Experimental Interactive Learning in Industrial Management of the IFIP Working Group. Milano, 2010. Retrieved October 15, 2010, from http://sites.google.com/site/ifip57sigexilim/

Thoben, K.-D., Baalsrud Hauge, J., Echelmeyer, W., & Morales, E. (2005). *Training through gaming: Applying a simulation based business game to train people for collaboration in virtual enterprises*. In Online Educa Berlin 2005, International Conference on Technology Supported Learning and Training.

UN. (1992). *United Nations agenda 21: Earth summit – The United Nations programme of action from Rio.* Retrieved October 15, 2010, from http://www.un.org/esa/dsd/agenda21/res_agenda21_00.shtml

UN. (2002). *United Nations report of the World Summit on Sustainable Development*. Retrieved October 15, 2010, from http://www.un.org/jsummit/html/documents/documents.html

WCED. (1987). *Our common future*. New York, NY: Oxford University Press.

WCSD. (2010). *E-brochure.* Retrieved October 15, 2010, from http://www.pagegangster.com/p/MwNhJ/

Wert, T. J. (2004). ICT-rich and competency based learning in higher education. In A. J. Kallenberg & M. J. J. M. van de Ven (Eds.), *The New Educational Benefits of ICT in Higher Education: Proceedings*. Rotterdam, The Netherlands: Erasmus Plus BV, OECR.

Westera, W. (2001). Competences in education: A confusion of tongues. *Journal of Curriculum Studies*, *33*(1), 75–88. doi:10.1080/00220270120625

ADDITIONAL READING

Beck, J. C., & Wade, M. (2005). *Got game. How the gamer generation is reshaping business forever*. Harvard Business School Press.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment. Exploring the educational potential of computer games*. Dissertation at the University of Copenhagen.

Gee, J. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave McMillan.

McClelland, D. C. (1973). Testing for competence rather than for "intelligence". *The American Psychologist*, *28*(1), 1–14. doi:10.1037/h0034092

Michael, D., & Chen, S. (2006). *Serious games: Games that educate, train and inform*. Boston, MA: Thomson Course Technology.

Spencer, L., & Spencer, S. (1993). *Competence at work: Models for superior performance*. New York, NY: John Wiley & Sons.

Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (2006). *Automated scoring of complex tasks in computer-based testing*. Lawrence Erlbaum Associates.

KEY TERMS AND DEFINITIONS

Competence: Within the TARGET project, a competence is defined as a specific, definable and measurable knowledge, skill, ability and/or other deployment-related characteristic (e.g. attitude, behavior, physical ability) which a human resource may possess and which is necessary for the performance of an activity within a work context. An overview of definitions of competences is available in Sampson and Fytros (2008).

Life Cycle Analysis: Is a "cradle-to-grave" approach for assessing industrial systems. "Cradle-to-grave" begins with the gathering of raw materials from the earth to create the product and ends at the point when all materials are returned to the earth. LCA evaluates all stages of a product's life from the perspective that they are interdependent, meaning that one operation leads to the next. LCA enables the estimation of the cumulative environmental impacts resulting from all stages in the product life cycle, often including impacts not considered in more traditional analyses (e.g., raw material extraction, material transportation, ultimate product disposal, etc.) (Curran, 2006).

Serious Gaming: Serious games are computer and/or video games used beside entertaining goals for educational technology. Serious games can be of any genre and many of them can be considered a kind of edutainment. Computer based serious games are an e-Learning methodology (Thoben et al, 2005)

Sustainable Development: Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own need (Brundtland, 1987).

Sustainable Manufacturing: Sustainable manufacturing deals with the application of sustainable development in the manufacturing sector. Commonly agreed measures in sustainable manufacturing include e.g. reduction of energy and raw material use, substitution of toxic and non-renewable materials, reduction of unwanted outputs (pollution, waste), reuse (recycling) of outputs, change structures of ownership in global manufacturing.

ENDNOTE

1 www.role-project.eu

This work was previously published in Handbook of Research on Serious Games as Educational, Business and Research Tools, edited by Maria Manuela Cruz-Cunha, pp. 47-71, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 44
Reengineering for Enterprise Resource Planning (ERP) Systems Implementation:
An Empirical Analysis of Assessing Critical Success Factors (CSFs) of Manufacturing Organizations

C. Annamalai
Universiti Sains Malaysia, Malaysia

T. Ramayah
Universiti Sains Malaysia, Malaysia

ABSTRACT

Reengineering is a concept that is applicable to all industries, particularly information and communication technology (ICT) projects regardless of organizational type, size, culture, or location. The enterprise resource planning (ERP) system frequently requires organizations to change their existing business processes to harmonize them its functional activities. 72% of the ERP implementation failures reported worldwide (Eric, 2010) because of the various critical success factors (CSFs). A Critical Success Factor (CSF) is defined as a factor needed to implement ERP system successfully. Assessing the importance of CSFs of Enterprise Resource Planning systems has always remained an important concern for academicians and researchers. This study explores and assesses the CSFs affecting the ERP implementation success. Long term Top management Support (LTS), Perceived ERP benefits (PEB), ERP in-house Training (EIT), Project Tracking (PTG), Visible Project Phases (VPP), Project Phase Update (PPU), Interdepartmental Cooperation (IDP), Strategic IT planning (STP), ERP vendor Support (EVS), and Data Analysis and Conversion (DAC) were found dominant critical factors for the success of the ERP implementation in the manufacturing sector. This study investigates how many CSFs are strongly correlated with each other for the success of ERP projects in the manufacturing sector. Furthermore, this study also tests empirically using the Statistical Package for Social Science Analysis of Moment on Structures (SPSS AMOS 18.0) to justify the level of CSFs among the local and joint-venture companies using a t-test analysis.

DOI: 10.4018/978-1-4666-1945-6.ch044

1 INTRODUCTION

An organization must proactively reengineer and plan for changes to business process before implementing a particular ERP module. The following are the important modules of ERP: sales and distribution, production planning, financials and controls, material management, and human resource management. Critical success factors (CSF) help implementing ERP system successfully. Alaranta (2006) pointed out the growth of information systems in the organizations resulted in the production of significant amounts of information for analysis and decision making which leads to the success of the information system projects, in particular enterprise resource planning (ERP) systems. Many research studies discussed numerous CSF factors that are needed to minimize the ERP failure rates (Eric, 2010; Cotran et al. 2005; Deloitte, 2005; Esteves, 2005; Al-Mashari, 2003; Nah & Lau, 2001).

A legacy system is an operational system that has been designed, implemented and installed in a radically different environment than that imposed by the current ICT strategy (Tromp & Hoffiman, 2008). The only reason a new system is developed is to replace an aging system (i.e. legacy system) that is failing to meet current enterprise needs. Ransom et al. (1998) pointed out that legacy systems are usually critical to the business in which they operate, but the costs of running them are often not justifiable. Furthermore, the legacy system contains the existing information technology (hardware and software), business processes, organization structure, and culture.

Appropriate business and legacy systems are important in the initiation stage of the project life cycle. Because of integrated nature of the ERP package, there is a choice to be made on the level of functionality and approach to link the system to legacy systems. In addition, to best meet business needs, companies may integrate other specialized software products (i.e. third party software packages or interfaces) with the ERP suite (Nah & Lau, 2001).

Many researchers highlighted that ERP implementation involves a complex transition from legacy information systems and business processes to an integrated IT infrastructure and common business process throughout the organisation (Al-Mudimigh et al. 2010; Jarrar et al. 2000; Gibson et al. 1999).

Al-Mashari (2003) and Seethamraju (1999) suggest that future ERP systems will be developed based on components rather than modules and will be designed for incremental migration rather than massive reengineering. Furthermore, Sato *et al*. (1999) identified several areas for future research, including integrating ERP and other business intelligence systems such as customer relationship management (CRM), supplier relationship management (SRM) and business data warehousing (BDS).

Most authors preferred an incremental approach to implementing either the business process reengineering (BPR) or ERP systems (Tromp & Hoffiman, 2008; Calvert, 2006; Robey et al. 2002). Hill (1994) pointed out rapid IT innovation and increasingly intensive global competition as two main reasons why organizations have had to consider the introduction of radical change. Reengineered processes drive the shape of an organization. These radical changes are not limited to inside the organization but can go beyond to other organizations, which generate innovative views for an organization (La Rock, 2003).

Most researchers lately found the actual acceptance of incremental or cyclical Enterprise Systems (ES) implementation approaches are slowly beginning to become authentic. In the last century the incremental implementation of ES was mentioned (Nagaraj et al. 2010; Mezeszaros & Aston, 2007; Karimi et al. 2007; Katsma & Spil, 2003). But the massive technology behind the ERP systems at first forbidden actual incremental implementation approaches. This technological hurdle is only slowly disappearing via the implementation of for example service oriented architecture (SOA) or software as a service (SAAS) based technologies.

The chapter is presented in ten main sections as follows: Section 2 describes research framework; Section 3 provides a detailed review of the CSFs, Section 4 discuss implementation success (IS), Section 5 illustrates data collection, Section 6 analyzing the results, Section 7 deals with the implications of the CSFs on IS, Section 8 discuss research limitations of this study, Section 9 describes future direction of this research and Section 10 concludes.

2 RESEARCH FRAMEWORK

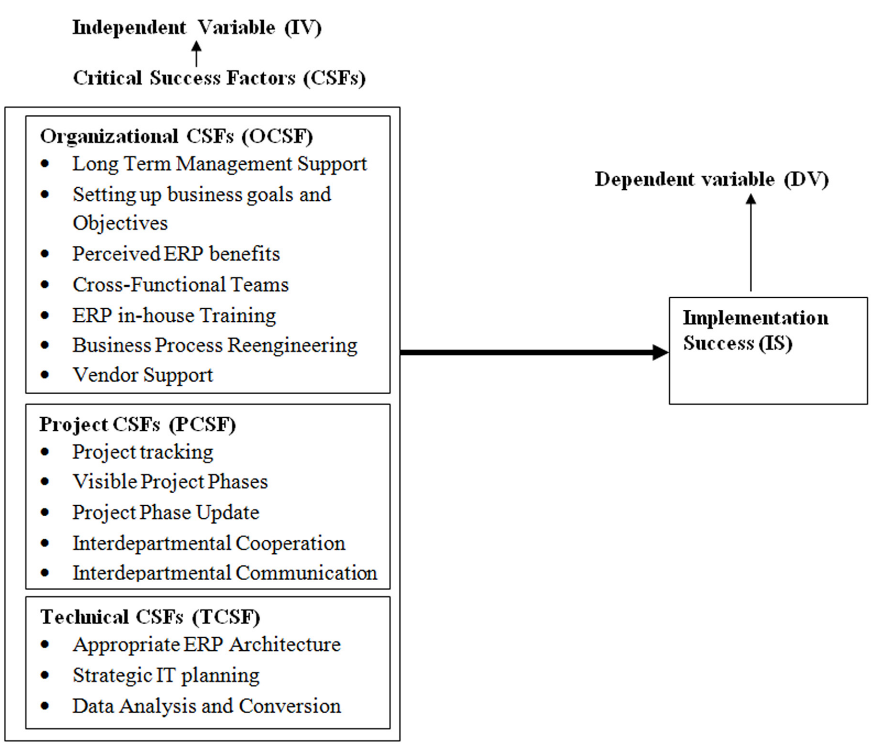
The research framework for the CSFs (the independent variable) on the ERP implementation success (IS as the dependent variable) is as shown in Figure 1. The above framework is proposed and used to assess the success of a migration effort from a legacy system to the new system that has to be measurable. Therefore, a number of critical success factors (CSF's) must be defined against the implementation success to measure the direct relationship between them. These success factors are used to check the correlations among each other which are very important to make a decision in the strategic plan and in line with the organization's business objectives (Tromp & Hoffiman, 2008). In this particular case, the following CSF's were identified based on the organizational, project and technological aspects which are described below in the next section.

3 LITERATURE REVIEW ON CSFs AND IMPLEMENTATION SUCCESS (IS)

3.1 Organizational Critical Success Factors (OCSF)

The Organizational Critical Success Factors consist of the following seven factors: long term

Figure 1. Proposed research framework



management support, setting up ERP business goals and objectives, perceived ERP benefits, cross-functional teams, ERP in-house training, business process reengineering and ERP vendor support.

3.1.1 Long Term Management Support (LTS)

Liang et al. (2007) explored that the top management support is the topmost CSF's among all critical success factors identified in the survey of information systems implementation success (IS). Furthermore, they argue that the entire team constitutes the senior and middle level management starting from the project initiation stage until the deployment stage in the entire ERP project life cycle (Carroll & Fitz-Gerald, 2005).

3.1.2 Setting up ERP Business Goals (SEB)

Remus (2006) and Geib et al. (2005) identified that a clear vision of business goals will help the ERP project better to achieve its expected business results. Furthermore, the value of information technology (IT) investments helps align between the business processes and strategies.

3.1.3 Perceived ERP Benefits (PEB)

Most researchers (Esteves, 2005; Freeman & Seddon, 2004; Gable et al. 2002) suggest that the employees need to encourage each other and work as a team in order to achieve the full potential benefits of ERP systems. O'Leary (2004) investigated the key benefits of ERP implementation and argues that the ERP tangible benefits are consistent across various industries such as manufacturing, hardware and software where as the ERP intangible benefits differ across the industry.

3.1.4 Cross-Functional Teams (CFT)

The communication must be shared across the cross-functional teams among the ERP with other systems; for instance Supplier Relationship Management (SRM) systems, Customer Relationship management (CRM) systems which will play a dominant role in the success of technology project. El Amrani et al. (2003) highlighted that this cross-functional systems help to reengineer and improve the business processes across the enterprise and further stated how challenging it was particularly in small medium enterprise (SME) sectors.

3.1.5 ERP In-House Training (EIT)

Mabert et al. (2003) and Sumner (2000) pointed out that when the ERP system is up and running it is very important that the users be capable of using, hence they should be aware of the ERP logic as well as concepts and should be familiar with the system's features.

3.1.6 Business Process Reengineering (BPR)

Markus and Tannis (2000) observed that the business process is based on best practices and reengineering the vision of business process through new methodology in information technology is crucial for information system success (Huq et al., 2006; Cao & Rahmati, 2005).

3.1.7 ERP Vendor Support (EVS)

Xue et al. (2005) and El Amrani et al. (2003) recognized that the ERP vendors need to provide support to the interfaces with cross-functional systems. In addition, they also have to realize the impact of implementation failures in cross-cultural issues.

3.2 Project Critical Success Factors (PCSF)

The Project Critical Success Factors consist of the following five factors: project tracking, visible project phases, project phase update, interdepartmental cooperation and interdepartmental communication.

3.2.1 Project Tracking (PTG)

Sumner (2006) pointed out that the project manager needs to track the projects from the initiation to the infusion stage with the respective team members in order to achieve better results.

3.2.2 Visible Project Phases (VPP)

Bajwa et al. (2004) and Esteves et al. (2003) studied the ERP project phases. They argue that it must be visible from the initiation to the infusion stage to the entire team members so that they can understand better and manage the whole project until deployment.

3.2.3 Project Phase Update (PPU)

According to Brazel (2005), the ERP project life cycle phases from the initiation to infusion stage need to be updated on a regular basis and to inform the team members as well as the top management in order to succeed the implementation of ERP systems. This is in consistent with Chang (2004) research study also.

3.2.4 Interdepartmental Cooperation (IDP)

Most projects have a collaborative requirement across other departments and conflicting schedules and constraints to generate the cooperation of the resources from other departments. The paradox being that the goal in introducing technology is to introduce business functions across several departments in a seamless fashion (Sternad & Bobek, 2006).

3.2.5 Interdepartmental Communication (IDC)

The importance of communication across different business functions and departments is well known in the IT implementation literature (Nah et al., 2001), because communication has a high impact from the initiation phase until the system acceptance, as it helps to minimize the user resistance. Communication has to cover the scope, objectives and tasks of an ERP implementation project. ERP systems need to have effective communication in project team and within the organization (Sternad & Bobek, 2006; Somers & Nelson, 2004).

3.3 Technical Critical Success Factors (TCSF)

The Technical Critical Success Factors consists of the following three factors: appropriate ERP architecture, Strategic IT planning and Data Analysis and Conversion. An appropriate ERP architecture should be in place before developing the ERP system (Molla & Loukis, 2005).

3.3.1 Appropriate ERP Architecture (AEA)

Holland and Light (2001) and Hung et al. (2004) argue that a good ERP architecture is well in place and should be discussed with the top management as well as other team members so that they can understand the entire model and its functions

3.3.2 Strategic IT Planning (STP)

Forcht et al. (2007) described that the strategic IT planning is an organization's process of defining its IT strategy, or direction, and making decisions on allocating its resources to pursue the strategy. Strategic IT planning also leads the ERP to achieve organization's competitive advantage. The process of strategic planning provides a means for ensuring that the entire organization is focused on a shared purpose and vision.

3.3.3 Data Analysis and Conversion (DAC)

The data analysis and conversion factor plays a vital role towards an implementation successful of an ERP system. If one may ignore problems with data that are not fixed in legacy systems, they will be apparent in the new system as well. ERP is an integrated package and all the modules in ERP are interlinked to one another; inaccurate data input into one module will adversely affect the functioning of other modules. (Jayaraman & Bhatti, 2007; Jafari et al., 2006; Akkermans & van Helden, 2002).

4 ERP IMPLEMENTATION SUCCESS (IS) AS A DEPENDENT VARIABLE

ERP implementation success, the dependent variable in this study, is different from the traditional information system success measure; which it is not evaluated by the general users but by the project team members. The term success means that the ERP project should be completed on time and within the budget (Chen et al. 2009; Magnusson et al. 2004; Hong & Kim 2002). ERP implementation success is measured in terms of the perceived deviation from the expected project goals such as cost overrun, schedule overrun, system performance deficit, and failure to achieve the following benefits: inventory reduction, personal reduction, productivity improvement, order management improvement, technology cost reduction, procurement cost reduction, cash management improvement, revenue/profit improvement, transportation/logistics cost reduction, maintenance reduction, and on time delivery improvement (Singla, 2008).

To sum up the literature review, the weakness of the above studies were just focusing only on organizational CSFs (OCSFs) and technical CSFs (TCSFs) but very few on the other hand about the importance of ERP systems' topmost project critical success factors (PCSFs) such as the visible project phase, the project tracking, and the project phase update. This gap needs to be considered especially when it comes to ERP implementation system because of its huge investments and its long-term benefits which would further lead to guidelines for practitioners and academicians.

5 DATA COLLECTION

The main goal of this study is to explore and identify the important Critical Success Factors (CSFs) for the successful ERP implementation in the manufacturing organization. In order to test the questionnaire of this study, it was first sent by email to the senior managers of ERP systems. After the feedback was processed and analysed, some minor changes were made to the wording and layout of the questionnaire.

The final copy of the research questionnaire was posted on the web page (i.e. Blog) at http://erpphd.blogspot.com/. A link to this URL was then sent to 1269 manufacturing organization. After several follow-up emails and phone calls, only 223 responses were received and all of them were used in the data analysis. Therefore, the response rate in this study is 20.82% when compared with other studies for ERP systems is higher than those represented by Nah et al. (2007) (5.4%), and Somers and Nelson (2001) (13.5%). The response to the survey indicates that there are sufficient observations to analyze the dataset with Structural Equation Model (SEM). Hair et al. (2006) consider that samples in the range of 150 to 400 are sufficient for SEM analysis.

6 RESULTS AND DATA ANALYSIS

Test of Significance on Difference of Means

Having analyzed the first section of the survey, it was found that some statistical tests would be helpful to further justify the level of CSFs among the local and joint-venture companies. Some tests are carried out to discover whether there was any significant difference between some aspects of CSFs factors on implementation success. These tests were analyzed using an ordinary comparison t-test. In order to conduct the tests, the following hypotheses were set up to test for a significant difference between the local and joint-venture companies' means on each of the CSFs factors:

H_0: There is no significant difference between the two means
H_1: There is significant difference between the two means

Significant Difference between the Local and Joint-Venture Companies

T-test analysis has been used in many research studies to compare different means of different groups (i.e. local and joint-venture companies). The t-test was carried out to investigate if there are any significance difference between the local and joint venture companies' means on each of CSFs. The result of the t-test is presented in Table 1.

Table 1 illustrated that the CSFs on IS were differed significantly ($p<.05$) with regards to their perceptions on the following CSFs: *Long term Top management Support (LTS), Perceived ERP benefits (PEB), ERP in-house Training (EIT), Project Tracking (PTG), Visible Project Phases (VPP), Project Phase Update (PPU), Interdepartmental Cooperation (IDP), Strategic IT planning (STP), ERP vendor Support (EVS), and Data Analysis and Conversion (DAC)*. These are the dominant CSFs which played a success role in ERP implementation in the manufacturing sector. Furthermore, it shows that there is a difference between the local and joint venture companies' CSFs towards an ERP implementation success in the manufacturing sector. When implementing the ERP projects, the implementers need to consider all the above CSFs and ultimately will discover that many more business processes would need reengineering after the ERP system went live.

The instrument measure of correlation analysis from the Table 2 shows that the *Long term Top management Support (LTS), Setting Business Goals and Objectives (SEB), Perceived ERP benefits (PEB), Business Process Reengineering (BPR), Visible Project Phase (VPP), Interdepartmental Cooperation (IDP), and Interdepartmental Communication (IDC*) are *strongly* correlated towards an ERP implementation success in the manufacturing sector. The CSFs such as *Cross-Functional Teams (CFT), Project Tracking (PTG), Project Phase Update (PPU), ERP vendor Support (EVS), and Data Analysis and Conversion (DAC)* have *medium level* correlation and the CSFs such as *ERP in-house Training (EIT), Appropriate ERP Architecture (AEA), and Strategic IT Planning (STP)* are *weakly* correlated toward an ERP implementation success in the manufacturing sector. The empirical result of the correlation analysis is given in Appendix 1.

7 IMPLICATIONS

The implications of this chapter include: literature review with altogether different viewpoint and synthesizing the CSFs for ERP implementation in the manufacturing sector. Such an implication provides the further scope of applying results from this review to successful ERP implementation. It is expected from the top management that CSFs would help them to better decision making in the ERP implementation from initiation to infusion of the ERP life cycle stages. While implementing

Table 1. T-test difference on CSFs means

Group Statistics					**Local and Joint-venture Results T-test**
CSFs	**Company Ownership**	**N (%)**	**Mean**	**STD**	**t-value Sig.**
LTS	1	47.1	4.153	0.933	3.747 0.000** (+)
	2	52.9	3.362	1.068	3.522 0.001
SEB	1	47.1	4.256	0.607	0.907 0.336 (-)
	2	52.9	4.120	0.872	0.771 0.445
BPR	1	47.1	4.452	0.339	5.870 0.000** (+)
	2	52.9	3.689	1.011	3.981 0.000
CFT	1	47.1	4.224	0.767	4.074 0.000** (+)
	2	52.9	3.482	1.004	3.605 0.001
PEB	1	47.1	4.554	0.479	8.522 0.000** (+)
	2	52.9	3.534	0.709	7.519 0.000
EIT	1	47.1	4.057	0.821	2.376 0.019** (+)
	2	52.9	3.551	1.318	1.932 0.061
PTG	1	47.1	4.493	0.543	4.980 0.000** (+)
	2	52.9	3.862	0.680	4.496 0.000
VPP	1	47.1	4.474	0.428	8.736 0.000** (+)
	2	52.9	3.528	0.651	7.253 0.000
PPU	1	47.1	4.397	0.591	5.116 0.000** (+)
	2	52.9	3.770	0.480	5.624 0.000
IDP	1	47.1	4.179	0.685	3.009 0.003** (+)
	2	52.9	3.586	1.338	2.279 0.029
IDC	1	47.1	4.282	0.664	0.040 0.968 (-)
	2	52.9	4.275	0.807	0.037 0.971
AEA	1	47.1	4.314	0.665	0.637 0.525 (-)
	2	52.9	4.206	1.013	0.529 0.600
DAC	1	47.1	4.352	0.684	2.410 0.018** (+)
	2	52.9	3.931	1.066	1.982 0.055
EVS	1	47.1	4.149	0.937	2.004 0.048** (+)
	2	52.9	3.701	1.245	1.762 0.086
STP	1	47.1	4.205	0.599	2.693 0.008** (+)
	2	52.9	3.793	0.930	2.219 0.033

**Significant $p<0.05$; 1: Joint-Venture; 2: Local; Hypothesis support: +:H_1 -:H_0

ERP projects, the reengineering experience have been difficult but rewarding to reap the tangible and intangible benefits.

8 LIMITATIONS

This chapter has some limitations. First, the study is limited to only Indian manufacturing organizations and cannot be generalized. Second, the study

Table 2. Results of CSFs correlation analysis

OCSF	Instrument Measure	Correlation
LTS	OCSF1.There have been top management support including commitment and leadership of management throughout ERP life cycle stages OCSF2.There have been enough knowledge on change management involving commitment to change and degree of process adaptation OCSF3.There have been role of top management communication about ERP benefits among all staffs OCSF4.There have been role of top management communications about ERP operations OCSF5. There have been top management support to allocate enough budget and resources	*Stronger*
SEB	OCSF6.There have been a clear goal and objectives of ERP business plan OCSF7. There have been set the goals of the ERP project before seeking top management support	*Stronger*
EIT	OCSF8. There have been training on how the new system will help improve business functions OCSF9. There have been training on how the ERP architecture works OCSF10. There have been training which should start early, preferably well before the implementation begins	*Weak*
BPR	OCSF11. There have been need to have a considerable knowledge in enterprises' Business Process Reengineering (BPR) methodology OCSF12. There have been alignment of BPR to ERP implementation OCSF13. There have been BPR which helps to contemporary measures such as cost, quality, service and speed	*Stronger*
EVS	OCSF14. There have been a good interpersonal skills between the organization and vendor OCSF15. There have been good vendor relationship with at least one firm for long term support after the implementation of the ERP project OCSF16. There have been vendors who are knowledgeable in enterprises' BPR methodology OCSF17. There have been Service Response Time which is almost immediate and lifelong commitment	*Medium*
CFT	OCSF18.There have been a policy to allow organizational team members to interface with inter-disciplinary team OCSF19. There have been mix of external consultants and internal staff involved in the ERP project OCSF20. There have been a policy that allows services and ERP functions without passing through hierarchical channels	*Medium*
PEB	OCSF21. ERP would improve in customer interaction and relationship OCSF22. ERP would improve supplier interaction and relationship OCSF23. ERP would considerably reduce raw material cost OCSF24. ERP would help to improve organization's strategic move OCSF25. ERP would facilitate better day-to-day management OCSF26. ERP would reduce lead time for customers OCSF27. ERP would considerably reduce production time OCSF28.ERP would improve organization's Return-On-Investments (ROI) OCSF29. ERP would help better resource management OCSF30. ERP would decrease Total Cost of Ownership (TCS)	*Stronger*
PCSF		
PTG	PCSF1. There have been empowered project manager and decision makers to track the ERP projects PCSF2. There have been a mechanism which would avoid budget overrun and ensuring implementation within schedule PCSF3. There have been a mechanism which would track the ERP project throughout the life cycle stages (i.e. Initiation Stage to Infusion Stage)	*Medium*
VPP	PCSF4.There have been visible project phases from the initiation to infusion stage of the project life cycle PCSF5. There have been a project champion who should provide strategic input for each project stage PCSF6. There have been a project champion market the ERP benefits back to the business process	*Stronger*
PPU	PCSF7. There have been a mechanism to update project phase throughout the ERP life cycle stages (i.e. from initiation to infusion stage) PCSF8. There have been a project champion who conduct project status update meetings periodically PCSF9. There have been a update process which should be continued with those of contractors or consultants, vendors and other parties involved in the implementation effort at each stage of ERP life cycle	*Medium*

continued on following page

Table 2. Continued

OCSF	Instrument Measure	Correlation
IDP	PCSF10. There have been cooperation of the various resources across other departments PCSF11. There have been cooperation which strongly influences the successful adoption and adaptation of enterprise resource systems PCSF12. There have been cooperation which helps to integrate all business processes within the organization	*Stronger*
IDC	PCSF13. There have been a communication which has a high impact from initiation phase until system acceptance PCSF14. There have been a communication which has to cover the scope, objectives and tasks of an ERP implementation project PCSF15. There have been a communication plan helps to minimize user resistance	
TCSF		
AEA	TCSF1. There have been appropriate business and IT legacy systems with right architectural choice TCSF2. There have been establishment of overall architectural planning prior to implementation TCSF3. There have been an appropriate architecture which should be based on the existing organizational requirements	*Weak*
DAC	TCSF4. There have been an ignoring problems in legacy systems which will affect the new system TCSF5. There have been the data residing in the legacy systems needs to be migrated to ERP system TCSF6. There have been software configuration and systems integration with completed testing TCSF7. There have been appropriate project methodology and vendor tools	*Medium*
STP	TCSF8. There have been a strategic IT planning which characterizes an organization's competence in matching IT capabilities TCSF9. There have been CSFs approach which has been used by managers as a framework for strategic planning to direct them and achieving the goals and objectives TCSF10. Strategic IT planning provides a means for ensuring that the entire organization is focused on a shared purpose and vision.	*Weak*

is based on the opinion of respondents (questionnaire) only. The questionnaire might have excluded some important factors. Third, the results will be different if one employs moderating variables such as organizational culture or implementation stage using ERP project life cycle stages. Fourth, the entire set of data analysis is based on three ERP packages namely SAP, ORACLE and RAMCO and cannot be generalized. Finally, the results were limited to Commercial-off-the-shelf (COTS) and customized ERP packages only.

9 FUTURE RESEARCH DIRECTIONS

This research was conducted among manufacturing sector, thus further research can be done with different sectors of the organizations and to make comparison between different types of organization. For example, it would be prudent to make comparison between government and private companies, as it will give rich information pertaining CSFs in both types of companies and whether there is a difference or not, in order to make a better strategic planning to implement ERP projects in the manufacturing sector.

10 CONCLUSION

This chapter was focused on the ERP critical success factors. Despite some limitations, this study is among the first to assess the CSFs of ERP implementation success, particularly in the manufacturing sector. The aim of this study was to investigate CSFs affecting the implementation success of the ERP projects in the manufacturing sector. Further, the proposed model explains that there is difference about the CSFs on the implementation success of the ERP projects by the local and joint-venture companies. There was a strong correlation among the *seven* out of the *fifteen critical success factors* identified towards an implementation success of ERP projects in the manufacturing sector. The ERP project is not a matter of changing software systems; rather it is a matter of repositioning the company and transforming business reengineering practices. If one choose customization of the software it results in higher implementation cost and longer duration of implementation.

REFERENCES

Akkermans, H., & van Helden, K. (2002). Vicious & virtuous cycles in ERP implementation: A case study of interrelations between critical success factors. *European Journal of Information Systems*, *11*, 35–46. doi:10.1057/palgrave/ejis/3000418

Al-Mashari, M. (2003). Enterprise resource planning (ERP) systems: A research agenda. *Industrial Management & Data Systems*, *103*(1), 22–27. doi:10.1108/02635570310456869

Al-Mudimigh, A., Zairi, M., & Al-Mashari, A. (2010). ERP implementation: An integrative methodology. *Proceedings of the European Centre for Best Business Practice Management, UK.*

Alaranta, M. (2006). Managing the change process of the post-merger enterprise systems integration: A case study. *International Journal of Information Systems and Change Management*, *1*(1), 62–81. doi:10.1504/IJISCM.2006.008287

Bajwa, D. S., Mooney, T., & Garcia, J. E. (2004). An integrative framework for the assimilation of enterprise resource planning systems: Phases, antecedents and outcomes. *Journal of Computer Information Systems*, *44*(3), 81–90.

Brazel, J. F. (2005). A measure of perceived auditor ERP systems expertise: Development, assessment, and uses. *Managerial Auditing Journal*, *20*(6), 619–631. doi:10.1108/02686900510606100

Calvert, C. (2006). A change-management model for the implementation and upgrade of ERP systems. *Proceedings in the Americas Conference on Information Systems* (*AMCIS 2006), Paper 18*. Retrieved December 29, 2010, from http://aisel.aisnet.org/acis2006/18

Cao, G., & Rahmati, N. (2005). BPR and ERP: A Chinese case. *Proceedings of the 16th Australasian Conference on Information Systems (ACIS),* Manly, NSW, Australia, vol. 19 (pp. 1-10).

Carroll, J., & Fitzgerald, L. (2005). Beyond an IT project: Governance in ERP-driven business transformation. *Proceedings of the Sixteenth Australasian Conference on Information Systems (ACIS 2005). University of Technology, Sydney*, (pp. 1-9).

Chang, S.-I. (2004). ERP life cycle implementation, management and support: Implications for practice and research. *Proceedings of the 37th Hawaii International Conference on System Sciences,* Waikoloa, Big Highland, Hawaii. Retrieved 11 April, 2010, from http://www2.computer.org/portal/web/csdl/abs/proceedings/hicss/2004/2056/08/205680227cabs.htm doi: 10.1109/HICSS.2004.1265556

Chen, C.-C., Law, C.-C. H., & Yang, S. C. (2009). Managing ERP implementation failure: A project management perspective. *Journal of Transactions on Engineering Management*, *50*(1), 157–170. doi:10.1109/TEM.2008.2009802

Computers in Human Behavior, 26(5), 1136-1148. doi: 10.1016/j.chb.2010.03.020

Cotran, K., Buchmeiser, U., Seguin, J.-L., & Pelster, B. (2005). HR's role in implementing JTI's global ERP system. *Strategic HR Review*, *4*(5), 24–27.

Deloitte. (2005, April). *Calling a change in the outsourcing market*. Retrieved April 14, 2010, from http://www.deloitte.com/dtt/cda/doc/content/us_outsourcing_callingachange.pdf

El Amrani, R., Rowe, F., & Geffroy-Maronnat, B. (2006). The effects of enterprise resource planning implementation strategy on cross-functionality. *Information Systems Journal*, *16*, 79–104. doi:10.1111/j.1365-2575.2006.00206.x

Eric, K. (2010). *ERP failures*. Retrieved December 4, 2010, from http://it.toolbox.com/blogs/erp-roi/one-key-reason-why-72-of-erp-system-implementations-fail-37200

Esteves, J. (2005). Addressing the justification of enterprise systems benefits: A desires and expectancy disconfirmation model. *Proceedings of the Eleventh Americas Conference on Information Systems,* Omaha (pp. 2079-2083). Retrieved March 16, 2010, from http://aisel.aisnet.org/amcis2005/125/

Esteves, J., Chan, R., Pastor, J., & Rosemann, M. (2003). An exploratory study of knowledge types relevance along enterprise systems implementation phases. *Proceedings of the 4th European Conference on Organizational Knowledge and Learning Capabilities,* Barcelona, (pp. 1-15). Retrieved April 22, 2010, from http://profesores.ie.edu/jmesteves/OKLC_2003_paper.pdf

Forcht, K. A., Aldridge, A., & Shorter, J. D. (2007). Implementing enterprise resource planning (ERP) for strategic competitive advantage. *Journal of Issues in Information Systems*, *8*(2), 425–429.

Freeman, P., & Seddon, P. (2004). Factors affecting the realisation of benefits from CRM packaged software-based work systems. *Proceedings of the Pacific Asia Conference on Information Systems,* Shanghai, China (pp. 203-214).

Gable, G., Timbrell, G., Sauer, C., & Chan, T. (2002). An examination of barriers to benefits-realisation from enterprise systems in the public service. Working Paper submitted to *European Conference on Information Systems,* (pp. 342-250). Retrieved June 22, 2010, from http://eprints.qut.edu.au/4745/1/4745.pdf

Geib, M., Kolbe, L., & Brenner, W. (2005). Customer relationship management in business networks: Lessons from the financial services industry in Germany and Switzerland. *Management Information Systems Quarterly*, *4*(1), 247–260.

Gibson, N., Holland, C., & Light, B. (1999, June). A case study of a fast track SAP R/3 implementation at Guilbert. *Electronic Markets*, 190–193.

Gupta, A. (2000). Enterprise resource planning: the emerging organizational value systems. *Industrial Management & Data Systems*, *100*(3), 114–118. doi:10.1108/02635570010286131

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2006). *Multivariate data analysis* (7th ed.). Pearson Education.

Hill, S. C. (1995). *A concise guide to the IDEF0 technique: A practical technique for business process reengineering*. Puyallup, WA: Enterprise Technology Concepts Inc.

Holland, C., & Light, B. (2001). A stage maturity model for enterprise resource planning systems use. *Journal of the Database for Advances in Information Systems*, *32*(2), 34–45.

Hong, K., & Kim, Y. G. (2002). The critical success factors for ERP implementation: An organizational fit perspective. *Journal of International Management*, *40*, 25–40.

Huang, S., Chang, I., Li, S., & Lin, M. (2004). Assessing risk in ERP projects: Identify and prioritize the factors. *Journal of Industrial Management and Data Systems*, *104*(8), 681–688. doi:10.1108/02635570410561672

Huq, Z. (2006). Six-Sigma implementation through competency based perspective (CBP). *Journal of Change Management*, *6*(3), 277–289. doi:10.1080/14697010600941825

Ifinedo, P., Rapp, B., Ifinedo, A., & Sundberg, K. (2010). Relationships among ERP post-implementation success constructs: An analysis at the organizational level.

Jafari, S. M., Osman, M. R., Yusuff, R. M., & Tang, S. H. (2006). ERP systems implementation in Malaysia: The importance of critical success factors. *International Journal of Engineering and Technology*, *3*(1), 125–131.

Jarrar, Y., Al-Mudimigh, A., & Zairi, M. (2000). ERP implementation critical success factors: The role and impact of business process management. *Proceeding of the 2000 IEEE International Conference on Management of Innovation and Technology*, 12-15 November, Singapore.

Jayaraman, V., & Bhatti, T. (2007). An empirical validation of the critical success factors for the implementation of ERP systems. *Proceedings of European and Mediterranean Conference on Information Systems.* Retrieved July 19, 2010, from http://www.iseing.org/emcis/EMCIS2008/Proceedings/Late%20Breaking%20Papers-Unrefereed/Contributions/LBP%206/EMCIS2008_TBERP.pdf

Karimi, J., Somers, T., & Bhattacherjee, A. (2007). The impact of ERP implementation on business process outcomes: A factor-based study. *Journal of Management Information Systems*, *24*(1), 101–134. doi:10.2753/MIS0742-1222240103

Katerattanakul, P., Hong, S., & Lee, J. (2006). Enterprise resource planning survey of Korean manufacturing firms *Management Research News, 29*(12), 820-837.

Katsma, C. P., & Spil, T. A. M. (2003). Aligning ambitions with project approaches in cyclical enterprise system implementations. *Proceedings in the Americas Conference on Information Systems* (AMCIS 2009). Retrieved December 29, 2010, from http://aisel.aisnet.org/amcis2009/110

La Rock, N. (2003). *Examining the relationship between business process reengineering and Information Technology*. MSc thesis. Bowie State University, Maryland, Europe.

Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *Management Information Systems Quarterly*, *31*(1), 59–87.

Mabert, V. A., Soni, A., & Venkataramanan, M. A. (2003). Enterprise resource planning: Managing the implementation process. *European Journal of Operational Research*, *146*(2), 302–314. doi:10.1016/S0377-2217(02)00551-9

Magnusson, J., Nilsson, A., & Carlsson, F. (2004). *Forecasting ERP implementation success: Towards a grounded framework*. Retrieved July 12, 2010, from http://is2.lse.ac.uk/asp/aspecis/20040100.pdf

Markus, M. L., & Tanis, C. (2000). The enterprise systems experience - From adoption to success. In Zmud, W. (Ed.), *Framing the domains of IT research: Glimpsing the future through the past* (pp. 173–207). Cincinnati, OH: Pinnaflex Educational Resources.

Mezeszaros, G., & Aston, J. (2007). Agile ERP: “You don’t know what you’ve got ‘till it’s gone!” *Agile Conference 2007,* (pp. 143-149).

Molla, A., & Loukis, K. (2005). *Success and failure of ERP technology transfer: A Framework for analysing congruence of host and system cultures*. Working paper series No. 24. Retrieved 14 June, 2010, from http://www.sed.manchester.ac.uk/idpm/research/publications/wp/di/documents/di_wp24.pdf

Nagaraj, S., Ramachandra, M., & Ratna Kumar, J. (2010). Cyclic approach to Web based project management. *International Journal of Computers and Applications*, *8*(5), 26–30. doi:10.5120/1205-1727

Nah, F. F.-H., Islam, Z., & Tan, M. (2007). Empirical assessment of factors influencing success of enterprise resource planning implementations. *Journal of Database Management*, *18*(4), 26–50. doi:10.4018/jdm.2007100102

Nah, F. F.-H., Lau, J. L.-S., & Kuang, J. (2001). Critical factors for successful implementation of enterprise systems. *Business Process Management Journal*, *7*(3), 285–296. doi:10.1108/14637150110392782

O'Leary, D. E. (2004). Enterprise resource planning (ERP) systems: An empirical analysis of benefits. *Journal of Emerging Technologies in Accounting*, *2*, 63–72. doi:10.2308/jeta.2004.1.1.63

Ransom, J., Sommerville, I., & Warren, I. (1998). A method for assessing legacy systems for evolution. *Proceedings of the 2nd Euromicro Conference on Software Maintenance and Reengineering* (CSMR'98). Retrieved 2 January, 2011, from http://docs.google.com/viewer?a=v&q=cache:n9A7n5QQ1AMJ:citeseerx.ist.psu.edu/viewdoc/download%3Fdoi%3D10.1.1.128.9889%26rep%3Drep1%26type%3Dpdf+A+Method+for+Assessing+Legacy+Systems+for+Evolution&hl=en&gl=my&pid=bl&srcid=ADGEESjUih6_0PbR-bW-C8nvjY0wYsPhkGf0OaAAWMcaZXgn9hN4M_PxRs-dazb4jcFbbl8X_CbkL3IJd4kwDOfHxi-0J68TUG1o4iZMOgaG6Q_IPuYbLm4CMlxA-ZUNf_B80Pb4v4wtIhr&sig=AHIEtbR_hcQIN-XNJ4OX2xV5AJcBsvW-3GQ

Remus, U. (2006). Critical success factors of implementing enterprise portals. *Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS-39),* Kauai, Hawaii, Paper 8, (p. 182a). doi: 10.1109/HICSS.2006.103

Robey, D., Ross, J. W., & Boudreau, M.-C. (2002). Learning to implement enterprise systems: An exploratory study of the dialectics of change. *Journal of Management Information Systems*, *19*(1), 17–46.

Robinson, L., & Hill, C. (1994). *Concise guide to the IDEFO technique: A practical approach to business process reengineering*. Enterprise Technology Concepts Inc.

Sato, R., Hotaka, R., & Noguchi, H. (1999). ERP and business process engineering in education. *Proceedings of the 3rd Annual SAP Asia Pacific Conference* (pp. 127-133).

Seethamraju, R. (1999). ERP and business management ± new directions for education and research. *Proceedings of the 3rd Annual SAP Asia Pacific Conference* (pp. 133-139).

Singla, A. R. (2008). Impact of ERP systems on small and midsized public sector enterprises. *International Journal of Theoretical and Applied Information Technology*, *4*(2), 119–131.

Somers, T. M., & Nelson, K. G. (2001). The impact of critical success factors across the stages of enterprise resource planning implementations. *Proceedings of the 34th Hawaii International Conference on System Sciences,* Maui, Hawaii (pp. 1-10). doi:10.1109/HICSS.2001.927129

Somers, T. M., & Nelson, K. G. (2004). Taxonomy of players and activities across the ERP project life cycle. *Information & Management*, *41*(3), 257–278. doi:10.1016/S0378-7206(03)00023-5

Sternad, S., & Bobek, S. (2006). Factors which have fatal influence on ERP implementation on Slovenian organizations. *Journal of Information and Organizational Sciences*, *30*(2), 279–293.

Sumner, M. (2006). *Enterprise resource planning* (1st ed.). New Delhi, India: Dorling Kindersley.

Tromp, H., & Hoffiman, G. (2008). *Evolution of legacy systems: Strategic and technological issues, based on a case study.* Retrieved 2 January, 2011, from http://soft.vub.ac.be/FFSE/Workshops/ELISA-submissions/11-TrompHoffman-position.pdf

Xue, Y., Liang, H., Boulton, W. R., & Snyder, C. A. (2005). ERP implementation failures in China: Case studies with implications for ERP vendors. *International Journal of Production Economics*, *97*(3), 279–295. doi:10.1016/j.ijpe.2004.07.008

KEY TERMS AND DEFINITIONS

Business Process Reengineering (BPR): The project process and activities have to be aligned with the new system.

Critical Success Factors (CSF): Critical Success Factor (CSF) is defined as the factors that are needed to implement the ERP system successfully.

Enterprise Resource Planning (ERP): Enterprise Resource Planning is an integrated package with a central database supported business intelligence application such as sales and distribution, service applications, financial management, manufacturing and inventory applications.

ERP In-House Training (EIT): Train potential users on how the architecture works and how new functionality relates to the business process of a new system.

Long-Term Top Management Support (LTS): The support from the top management is important throughout the project life cycle in order to conduct a successful ERP implementation.

Perceived ERP Benefits (PEB): The primary benefits that are expected to result from the ERP implementation are closely related to the level of integration that is promoted across functions in an enterprise.

Project Phase Update (PPU): Project phases are spread out thought ERP project life cycle and are critical from the initiation phase to infusion stage and keep updated at regular intervals.

Project Tracking (PTG): Tracking the ERP projects from all phases of ERP life cycle (i.e. initiation to infusion stage).

Setting up ERP Business Goals (SEB): It is the objectives that are specific to the scope of the corporate project, the user community that is affected, and the timeline that needs to be met.

SPSS AMOS: Statistical Package for Social Science Analysis of Moments on Structures.

T-Test: The t-test assesses whether the means of two groups are statistically different from each other.

Visible Project Phases (VPP): It is the individual project stage from the initiation to infusion in an ERP project life cycle which must be visible to all team members.

APPENDIX: CSFS CORRELATION ANALYSIS

Correlation Analysis			
CSFs	**Company Ownership**	**N (%)**	**Correlation**
LTS	1	47.1	.746
	2	52.9	.769
SEB	1	47.1	.722
	2	52.9	.712
BPR	1	47.1	.622
	2	52.9	.652
CFT	1	47.1	.521
	2	52.9	.543
PEB	1	47.1	.699
	2	52.9	.612
EIT	1	47.1	.329
	2	52.9	.321
PTG	1	47.1	.501
	2	52.9	.518
VPP	1	47.1	.619
	2	52.9	.627
PPU	1	47.1	.543
	2	52.9	.562
IDP	1	47.1	.690
	2	52.9	.710
IDC	1	47.1	.677
	2	52.9	.719
AEA	1	47.1	.341
	2	52.9	.370
DAC	1	47.1	.497
	2	52.9	.519
EVS	1	47.1	.443
	2	52.9	.459
STP	1	47.1	.310
	2	52.9	.339

This work was previously published in Information Systems Reengineering for Modern Business Systems: ERP, Supply Chain and E-Commerce Management Solutions, edited by Raul Valverde and Malleswara Rao Talla, pp. 185-200, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 45
Optimal Pricing and Inventory Decisions for Fashion Retailers under Value-At-Risk Objective:
Applications and Review

Chun-Hung Chiu
City University of Hong Kong, Hong Kong

Jin-Hui Zheng
The Hong Kong Polytechnic University, Hong Kong

Tsan-Ming Choi
The Hong Kong Polytechnic University, Hong Kong

ABSTRACT

Motivated by the popularity of Value-at-Risk (VaR) objective in finance, this chapter reviews and studies its application in fashion retail operations management. First, a formal optimization model is reviewed in which the fashion retailer aims at optimizing a VaR objective with both pricing and inventory decisions. Second, the detailed solution schemes are explored. Third, numerical examples are included to illustrate applications of the proposed models. Fourth, the performance of pure buyback contract and pure wholesale pricing contract in enhancing supply chain efficiency is examined. Insights are generated and future research directions are outlined.

INTRODUCTION

The fashion retailing problem and the newsvendor problem share many common features. The fashion retailer faces uncertain demands in selling fashion products. The purchase cost of a specific fashion product is fixed, the selling season is short, the end-of-season markdown will clear the unsold leftover, and the fashion retailer needs to determine the order quantity of the fashion product from his supplier before the selling season starts. If we check the classical newsvendor problem (see, e.g., Lau, 1980), we can see that the problem faced by

DOI: 10.4018/978-1-4666-1945-6.ch045

a fashion retailer is very close to the newsvendor problem. As a result, there are quite a number of studies which employ the newsvendor model to investigate supply chain management problems in fashion retailing. Examples include an investigation of trade promotion scan-back rebate scheme in fashion supply chains (Kurata & Yue, 2008), a mean-variance analysis of quick response fashion supply chains (Choi & Chow, 2008), and a study on risk hedging problem in multi-product fashion supply chains (Vaagen & Wallace, 2008).

Over the past few years, there are extensive studies of inventory models with the consideration of risk in the literature (see, e.g., Wu et al., 2009; Tapiero, 2005; Choi and Ruszczyriski, 2008). However, most of these studies do not consider pricing and only explore the corresponding optimal inventory policies. In addition, for those which include both inventory and pricing decisions, they only illustrate the technical details but do not show the applications of the algorithms (e.g. Chiu & Choi 2010). Since retail pricing is undoubtedly a crucial decision for fashion retailers and applications are important, we explore in this chapter the joint optimal pricing and stocking algorithms with an optimization objective based on the Value-at-Risk (VaR) function for a single-period product with price-dependent demand. We propose detailed algorithms to illustrate how the optimal pricing and stocking policies can be determined for different forms of price-dependent demand functions. This chapter can be taken as a review plus extended works supplementing the related literature (such as Gotoh & Takano, 2007; Chiu & Choi, 2010) with detailed algorithms, numerical analysis and discussions on channel coordination challenge.

The rest of the chapter is organized as follows. A literature review is first provided. The basic model is then discussed. Afterwards, the solution schemes are proposed and numerical examples are provided to further illustrate the findings. The performance of pure buyback contract and pure wholesale pricing contract in enhancing supply chain efficiency is examined. We finally conclude with a discussion on probable future research directions.

LITERATURE REVIEW

As we mentioned earlier, the newsvendor problem, despite being simple, has been employed widely to study stochastic supply chain inventory problems for fashion products. One of the major differences between the fashion retailing problem and the classical newsvendor problem is that the fashion retailer can (and actually will) decide the retail price by his own while the selling price of newspapers is fixed for the newsvendor in the classical newsvendor problem. In the literature, there are several extensions investigating pricing issue for newsvendor problem. For example, Petruzzi and Dada (1999) review and extend the newsvendor model with price-dependent demands. Under the respective analytical models, they study the newsvendor problem in which the optimal stocking quantity and selling price are jointly determined. Important theoretical insights and structural properties of the "price-setting" newsvendors are hence developed. Later on, Dana and Petruzzi (2001) extend the price-dependent demand newsvendor problem by considering the expected utility maximizing consumers who would choose between the firm's product and the competitor's product. Other related works which make use of price-dependent newsvendor models include a study of the supply chain coordination problem between a newsvendor and its supplier with price-dependent demand (Chiu et al., 2010), and a study of newsvendor model which captures the joint pricing and procurement of fashion products in the existence of clearance markets (Karakul, 2008). All these works include the joint optimal decisions in both inventory and price, and they are known to be more challenging than the works that focus solely on the inventory (i.e., quantity) decision. In fact, notice that even for the expected profit maximizing newsvendor problem, the existence of a unique joint solution (of quantity and price) is not universally true for any distributions (see Petruzzi & Dada, 1999). Moreover, between the newsvendor problem and the fashion retailer's inventory and pricing problem, there exists another major difference in which the classical newsvendor model considers a fixed and known salvage value for the unsold products while this "salvage value" can vary or even be unknown for a fashion retailer (Ferdows et al., 2004). Even though we do not consider the salvage value issues in current chapter, some prior studies which extend the inventory decision model with investigation of this issue can be found (Cachon & Kok, 2007; Karakul, 2008).

In addition to the extensions with the price-dependent demand models, there are a lot of studies that extend the newsvendor problem in other directions. These extensions include the study of a newsvendor pricing game (Chen et al., 2004), the study of a capital constrained newsvendor problem (Dada & Hu, 2008), and the newsvendor problem research with resalable returns (Mostard & Teunter, 2006), etc. In particular, more and more studies related to the newsvendor problems explore risk related issues. For instance, pioneered by Lau (1980), the use of a mean-variance (MV) objective in the newsvendor setting has become popular in exploring risk averse newsvendors. Choi et al. (2008a) apply the MV analysis to study the newsvendor problems in which decision makers with different risk-attitudes are examined. Wu et al. (2009) perform MV analysis in the newsvendor setting with a stock-out cost and they demonstrate an interesting result that the optimal newsvendor quantity under the MV setting for a risk averse decision maker can be larger than that of the risk neutral case for a demand following a power distribution. Choi et al. (2008a) apply the MV analysis to the buyback contract in a two-stage supply chain. Besides the use of MV model, the VaR and conditional value-at-risk (CVaR) approaches have also been applied in the newsvendor problem and several inventory control models.

Some related works include: the incorporation of the CVaR and mean-CVaR in the newsvendor setting (Gotoh & Takano, 2007), the VaR approach in the multi-period inventory problem (Luciano et al., 2003), the multi-product newsvendor problem with VaR objective (Özler et al., 2008), optimal inventory control with VaR measures (Tapiero, 2005). Most recently, Chiu and Choi (2010) explore the structural properties and optimization problems of the newsvendor problem with VaR considerations. Although VaR (or CVaR) has been explored in the inventory related literature, the existing works seldom demonstrate the applications of the VaR approach in a relatively realistic setting. Plus, the issue of supply chain coordination with a supply chain agent which employs a VaR objective has not been explored. As a consequence, this chapter aims at reviewing some important results related to VaR objective in joint pricing and inventory problem, exploring the details of the respective applications, and examining the channel coordination issues. Numerical examples are included and future research directions are also discussed.

THE MODEL: A REVIEW

In this section we review an analytical model which basically follows the standard newsvendor problem with price-dependent demands. First, we consider a fashion retailer which sells a short-life fashionable product with the unit purchasing cost (i.e., product's wholesale price) c, and the unit retail selling price r. Any unsold product will be cleared with a big discount at "markdown price" and we denote the net unit value of the leftover product as v. Notice that in the classical newsvendor problem, v is usually assumed to be positive. However, for the fashion retailer, v can be negative especially for some fast fashion retailers of which product turnover rate is very high and usually leftover has negative effects on their brand images and promotion of new products (Ferdows et al., 2004).

The fashion retailer makes two important decisions before the season starts: the optimal retail price and the optimal order quantity. For the demand functions, two common forms of price-dependent demand functions, namely the linear demand and the multiplicative demand, are considered (Petruzzi & Dada, 1999; Chiu & Choi, 2010): (a) The linear (L) demand: $D_L(r)=k-ar+x$, *for* $0<r\leq k/a$, where $k>0$, $a>0$; and (b) The multiplicative (M) demand: $D_M(r)=gr^{-b}x$, where $g>0$ and $b>1$. For both linear demand and multiplicative demand, we consider x as a non-negative random variable with mean μ and standard deviation σ, with a density function $f(\cdot)$ and a cumulative distribution function $F(\cdot)$. We define the inverse function of $F(\cdot)$ by $F^{-1}(\cdot)$, and $\bar{F}(\cdot)=1-F(\cdot)$.

Noting that the above fashion retailing model is similar to the newsvendor model with price-dependent demand. Therefore, we can apply the price-dependent demand newsvendor problem to capture the fashion retailing problem. Now, we consider the case that the fashion retailer is risk-sensitive and it follows the VaR approach to determine the joint optimal retail price and order quantity. We first define $\Pi_i(q,r)$ as the fashion retailer's profit under demand function $i=L,M$. For a given confidence level α, where $0<\alpha<1$, suppose that a fashion retailer wants to determine the optimal quantity and selling price so that the following is achieved,

$$\Pr(\Pi_i(q,r)\leq V)=1-\alpha,$$

where V is the profit threshold, and the probability of having $\Pi_i(q,r)\leq V$ is $1-\alpha$. Notice that V is called the VaR, with a confidence level of α.

Now, in our optimization problem, the fashion retailer would like to maximize V by selecting the optimal price and quantity subject to a constraint that the probability of having $\Pi_i(q,r)\leq V$ is no larger than $1-\alpha$:

$$\max_{(q,r)} V$$

$$\text{s.t. } \Pr(\Pi_i(q,r)\leq V)\leq 1-\alpha.$$

Notice that the above optimization problem has been defined and employed in the literature (Chiu and Choi, 2010). For a notational purpose, we define $C(\alpha)=F^{-1}(1-\alpha)$.

A SOLUTION SCHEME

With the model reviewed and defined, we present the solution schemes in this section for both the linear and multiplicative price dependent demand functions. Based on the structural properties of the optimization problem defined above and the analytical results revealed in the literature (Chiu and Choi, 2010), we have the following two algorithms for any given confidence level $0<\alpha<1$,

For the Linear Price Dependent Demand - Algorithm L:

Step 1: Define and compute: $K_{L-}=k-ac$, $K_{L+}=k+ac$.
Step 2: Compute $C(\alpha)$.
Step 3: If $K_{L-}>C(\alpha)$: The optimal retail price = $0.5(C(\alpha)+K_{L+})a^{-1}$, the optimal quantity = $0.5(C(\alpha)+K_{L-})$, the corresponding optimal VaR = $0.25(C(\alpha)+K_{L-})a^{-1}$, and the corresponding service level = $1-\alpha$; otherwise, the optimal retail price = ka^{-1}, the optimal quantity = $C(\alpha)$, the corresponding optimal VaR = $C(\alpha)+K_{L-})a^{-1}$, and the corresponding service level = $1-\alpha$.

For the Multiplicative Price Dependent Demand - Algorithm M:

Step 1: Define and compute: $B=(b-1)$, $Z=bc$, and $\beta=B/Z$.
Step 2: Compute $C(\alpha)$.

Step 3: The optimal retail price = Z/B, the optimal quantity = $gC(\alpha)(B/Z)^b$, the corresponding optimal VaR = $gC(\alpha)b^{-1}$ $(B/Z)^{b-1}$, and the corresponding service level = $1-\alpha$.

NUMERICAL EXAMPLE

In this section, we present two numerical examples.

Example 1. In this numerical example, x follows a normal distribution with mean μ and standard deviation σ, and the market parameters are given in Table 1(a).

The optimal joint solution of retail pricing and stocking quantity (inventory) under the linear demand and the multiplicative demand, and the corresponding optimal VaR values for different α are shown in Table 1(b). Moreover, the service level and the stock out probability for different values are also shown in Table 1(b). Note that for the sake of simplicity, the salvage value v is set as zero in Table 1(b).

Table 1(a). Parameter setting of Example 1

c	μ	σ	a	k	B	g
100	100	10	0.5	200	1.5	10000

Notice that α and $VaR_{i,\alpha}$ have some managerial meanings for the fashion retailer. The fashion retailer with a VaR objective can make optimal decisions based on its own managerial requirements; for example, a VaR fashion retailer who wants to provide a 95% inventory service level to the customer should follow the row with α=0.05 in Table 1(b) in setting the respective parameters (i.e., set the optimal price = 300 and optimal quantity =224 under the multiplicative demand case). Another fashion retailer (with VaR objective and under the linear demand case) who wants to sell the fashion product at a price 340, should follow the row with α=0.85 in Table 1(b), i.e., set the optimal quantity=120, and bear a high sold out probability of 0.85 but only a small maximum attainable profit of 28713.

Example 2. In this numerical example, the focus is on the profits of the VaR fashion retailers with α = 0.05 (high service level), α = 0.55 (middle between high service level and high stock out probability) and α = 0.95 (high stock out probability), when x varies. Here, x follows a normal distribution, and the market parameters are given in Table 2(a).

The optimal joint solution of retail price and stocking quantity under the linear demand and multiplicative demand, and the corresponding

Table 1(b). Optimal retail price and stocking inventory of Example 1

α	Service level	Sold out probability	Linear Demand			Multiplicative Demand		
			Price	Quantity	*VaR*	Price	Quantity	*VaR*
0.05	0.95	0.05	366.45	133	35497.41	300.00	224	44821.06
0.15	0.85	0.15	360.36	130	33894.79	300.00	212	42479.25
0.25	0.75	0.25	356.74	128	32958.97	300.00	205	41086.13
0.35	0.65	0.35	353.85	127	32220.72	300.00	200	39973.12
0.45	0.55	0.45	351.26	126	31564.94	300.00	195	38973.69
0.55	0.45	0.55	348.74	124	30936.64	300.00	190	38006.35
0.65	0.35	0.65	346.15	123	30294.12	300.00	185	37006.92
0.75	0.25	0.75	343.26	122	29586.52	300.00	179	35893.91
0.85	0.15	0.85	339.64	120	28712.63	300.00	173	34500.78
0.95	0.05	0.95	333.55	117	27273.14	300.00	161	32158.97

Table 2(a). Values of market parameters of Example 5.2

c	μ	σ	a	k	b	g	v
100	100	30	0.5	200	1.5	10000	0

optimal VaRs with $\alpha = 0.05$, $\alpha = 0.55$ and $\alpha = 0.95$ are as shown in Table 2(b). The profit of the VaR fashion retailers having $\alpha = 0.05$, $\alpha = 0.55$ and $\alpha = 0.95$, under different market situation, i.e., different values of x, are summarized in Table 2(c).

Table 2(c) shows that the maximum attainable profit of the VaR fashion retailer is decreasing in α, while the probability of achieving the maximum profit is increasing in α. Moreover, when the market situation is bad, i.e., x is small, the profit of the VaR fashion retailer is increasing in α. For some cases, e.g., when both x and α are small, the VaR fashion retailer suffers a loss. Furthermore, the variation of profit under both the linear demand and the multiplicative demand cases for α =0.05 is the biggest among the three listed cases. Therefore, a small α of the VaR fashion retailer implies that the VaR fashion retailer shares (1) a high maximum attainable profit, (2) a low probability that achieves the maximum profit, and (3) a high variation of profit.

COORDINATION

We now consider a supply chain in which there are two echelons with one upstream manufacturer and a downstream retailer. The retailer takes the VaR objective (we call it "VaR retailer") and makes the optimal pricing and ordering quantity decisions accordingly. In the literature, under the expected profit maximizing model, owing to the double marginalization effect, a decentralized supply chain would not be optimal by itself. In particular, for the case when the retailer only makes the ordering quantity decision, a supply contract in the form of a pure wholesale pricing

Table 2(b). Optimal retail price and stocking inventory of Example 5.2

α	Service level	Sold out probability	Linear Demand			Multiplicative Demand		
			Price	Quantity	VaR	Price	Quantity	VaR
0.05	0.95	0.05	399.35	150	44803.90	300.00	287	57483.15
0.55	0.45	0.55	346.23	123	30314.65	300.00	185	37039.01
0.95	0.05	0.95	300.65	100	20131.09	300.00	97	19496.88

Table 2(c). Profit of VaR fashion retailers in Example 5.2

x	$F(x)$	Profit under LD			Profit under MD		
		α=**0.05**	α=**0.55**	α=**0.95**	α=**0.05**	α=**0.55**	α=**0.95**
20	0.0038	-6849.70	3921.47	10914.72	-17194.57	-6972.50	1798.56
40	0.0228	1137.21	10846.07	16927.80	-5647.57	4574.51	13345.57
60	0.0912	9124.12	17770.67	20131.09	5899.44	16121.51	19496.88
80	0.2525	17111.03	24695.28	20131.09	17446.45	27668.52	19496.88
100	0.5000	25097.94	30314.65	20131.09	28993.45	37039.01	19496.88
120	0.7475	33084.86	30314.65	20131.09	40540.46	37039.01	19496.88
140	0.9088	41071.77	30314.65	20131.09	52087.46	37039.01	19496.88
160	0.9772	44803.90	30314.65	20131.09	57483.15	37039.01	19496.88

contract (PWPC) cannot lead to an optimal supply chain (i.e., cannot coordinate the supply chain) whereas a pure buyback contract (PBC) can. For the case when the retailer makes both the retail pricing and ordering quantity decisions, both pure PWPC and PBC will fail to coordinate the supply chain even though PBC can enhance the supply chain's efficiency. In the following, we compare the performance of PBC and PWPC in enhancing the supply chain's efficiency with a VaR fashion retailer (as we modelled in the earlier sections).

First, we represent the unit production cost of the product by p, and the optimal supply chain price and product quantity are denoted by q_{SC*} and r_{SC*}, respectively. Under PBC, the manufacturer promises to buyback any unsold quantity by the retailer and the respective unit buyback price is denoted by E. Second, in order to guarantee the existence and uniqueness of q_{SC*} and r_{SC*}, we need some mild conditions on the demand distribution. To be specific, we assume that $\forall x \geq 0$: (a) $F(x)$ has a unique inverse function, and (b)$(1-F(x))^2/f(x)$ is decreasing in x (Petruzzi and Dada, 1999). For the multiplicative case, the logic is similar and hence we omit the details here.

We proceed to explore the performance of PWPC. First, we have Observation 1.

Observation 1: *Under PWPC: (a) For the multiplicative demand function: The supply chain can be coordinated if and only if* $q_{SC*}=gC(\alpha)[(b-1)/bc]^b$ and $r_{SC*}=bc/(b-1)$. (b) For the linear demand function: (i) When $c<(k-C(\alpha))/a$: The supply chain can be coordinated if and only if $q_{SC*}=0.5(C(\alpha)+k-ac)$ and $r_{SC*}=0.5(C(\alpha)+k+ac)a^{-1}$. (ii) When $c\geq(k-C(\alpha))/a$: The supply chain can be coordinated if and only if $q_{SC*}=C(\alpha)$ and $r_{SC*}=ka^{-1}$.

Proof of Observation 1: First, notice that by definition, coordination means providing incentive to make the retailer's optimal decisions under VaR objective to be the same as the supply chain's best decisions. Second, following Algorithms L and M, we can find the closed form expressions for the optimal decisions of the retailer under different demand functions. Combining the above two points yield the necessary and sufficient conditions and complete the proof. (Q.E.D.)

Obviously, *the required necessary and sufficient conditions as summarized by Observation 1 are unlikely to be satisfied in a general supply chain*. In other words, PWPC probably fails to coordinate the supply chain with price-dependent demands but in most cases (the multiplicative demand case, and case (b-i) of the linear demand), the retailer's optimal decisions do depend on the wholesale price c and hence adjusting c can enhance the supply chain's efficiency by making the retailer's optimal decisions closer to q_{SC*} and r_{SC*}. We summarize the key findings in Proposition 1:

Proposition 1: *In general, PWPC cannot coordinate the supply chain with a VaR retailer but it can enhance the supply chain's efficiency by making the optimal pricing and ordering decisions of the VaR fashion retailer closer to the optimal decisions of the supply chain.*

Suppose that we consider a scenario in which there is a fixed retail selling price policy in the supply chain, i.e., the retailer sells the product at a fixed retail price (which is exogenously given). This case arises in several situations, e.g., when there is a recommended standard price as requested by the fashion brand owner (for branding positioning reason), etc. In this case, the retailer only makes a decision on the optimal ordering quantity and coordination becomes achievable by the PWPC. Proposition 2 summarizes the result and the proof is similar to the proof of Observation 1.

Proposition 2: *Under PWPC and the retail selling price is exogenous: (a) For the multiplicative demand function: The supply chain can be coordinated if and only if* $c=(b-1)[gC(\alpha)/q_{SC*}]^{1/b}$. (b) For the linear demand function: The supply chain can be coordinated by settling the wholesale c such that $c=(C(\alpha)+k-2q_{SC*})/a$ and $c<(k-C(\alpha))/a$.

Following the same logic, we proceed to explore PBC. From Algorithms L and M, it is easy to prove Proposition 3:

Proposition 3: *PBC has zero impact on the optimal pricing and ordering decisions of the fashion retailer under the VaR objective. As a result, it is an ineffective incentive alignment scheme for enhancing supply chain's efficiency.*

Proposition 3 shows an interesting result that counter-intuitively, PBC can neither coordinate the supply chain with a VaR fashion retailer, nor enhance the supply chain's efficiency. This result shows a big challenge to the existing literature on the effectiveness of PBC and calls for more research to investigate the performance of different supply contracts in coordinating the supply chain when individual agents take different optimization objectives (Choi et al., 2008b).

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

VaR is a popular performance measure/objective and a hot topic in risk management. Based on the existing literature, this chapter reviews and studies VaR's application in deriving the optimal pricing and inventory decisions in fashion retailing. We first review the standard analytical model and discuss the VaR related optimization model for the fashion retailer. We then propose two detailed solution schemes, one for the linear price dependent demand and one for the multiplicative price dependent demand. Numerical examples are included to illustrate applications of the proposed models and to reveal some properties of the solutions. For future research directions, one can extend the work by further considering the effect of advertising with "advertising-effort and price" dependent consumer demands and examine the respective optimal decisions with a VaR objective. Following our analysis on channel coordination, one can investigate whether some hybrid policies can successfully achieve channel coordination when the fashion retailer takes a VaR objective. It is interesting to explore the joint pricing and stocking inventory decisions for the fashion retailer with a VaR objective in both the multi-period and multi-product cases. This extension is challenging while deserves deep investigation.

ACKNOWLEDGMENT

We sincerely thank the three anonymous reviewers for their comments which help enhance the quality of this chapter. This work was supported in part by the Hong Kong Research Grants Council Competitive Earmarked Research Grant under a grant number of PolyU 5143/07E.

REFERENCES

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*(3), 203–228. doi:10.1111/1467-9965.00068

Cachon, G. P., & Kok, A. G. (2007). Implementation of the newsvendor model with clearance pricing: How to (and how not to) estimate a salvage value. *Manufacturing and Service Operations Management*, *9*(3), 276–290. doi:10.1287/msom.1060.0145

Chen, F. Y., Yan, H., & Yao, L. (2004). A newsvendor pricing game. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *34*(4), 450–456. doi:10.1109/TSMCA.2004.826290

Chiu, C. H., & Choi, T. M. (2010). Optimal pricing and stocking decisions for newsvendor problem with value-at-risk (VaR) consideration. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *40*(5), 1116–1119. doi:10.1109/TSMCA.2010.2043947

Chiu, C. H., Choi, T. M., & Tang, C. S. (2010). Price, rebate, and returns supply contracts for coordinating supply chains with price dependent demand. *Production and Operations Management*, *20*(1), 81–91. doi:10.1111/j.1937-5956.2010.01159.x

Choi, S., & Ruszczyński, A. (2008). A risk-averse newsvendor with law invariant coherent measures of risk. *Operations Research Letters*, *36*(1), 77–82. doi:10.1016/j.orl.2007.04.008

Choi, T. M., & Chow, P. S. (2008). Mean-variance analysis of quick response program. *International Journal of Production Economics*, *114*(2), 456–475. doi:10.1016/j.ijpe.2007.06.009

Choi, T. M., Li, D., & Yan, H. (2008a). Mean-variance analysis for the newsvendor problem. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *38*(5), 1169–1180. doi:10.1109/TSMCA.2008.2001057

Choi, T. M., Li, D., Yan, H., & Chiu, C. H. (2008b). Channel coordination in supply chains with agents having mean-variance objectives. *Omega*, *36*(4), 565–576. doi:10.1016/j.omega.2006.12.003

Dada, M., & Hu, Q. (2008). Financing newsvendor inventory. *Operations Research Letters*, *36*(5), 569–573. doi:10.1016/j.orl.2008.06.004

Dana, J. D., & Petruzzi, N. C. (2001). Note: The newsvendor model with endogenous demand. *Management Science*, *47*(11), 1488–1497. doi:10.1287/mnsc.47.11.1488.10252

Ferdows, K., Lewis, M. A., & Machuca, J. A. D. (2004). Rapid-fire fulfillment. *Harvard Business Review*, *82*(11), 104–110.

Gotoh, J., & Takano, Y. (2007). Newsvendor solutions via conditional value-at-risk minimization. *European Journal of Operational Research*, *179*(1), 80–96. doi:10.1016/j.ejor.2006.03.022

Karakul, M. (2008). Joint pricing and procurement of fashion products in the existence of clearance markets. *International Journal of Production Economics*, *144*(2), 487–506. doi:10.1016/j.ijpe.2007.03.026

Kurata, H., & Yue, X. (2008). Trade promotion mode choice and information sharing in fashion retail supply chains. *International Journal of Production Economics*, *114*(2), 507–519. doi:10.1016/j.ijpe.2007.05.021

Lau, H. S. (1980). The newsboy problem under alternative optimization objectives. *The Journal of the Operational Research Society*, *31*(6), 525–535.

Luciano, E., Peccati, L., & Cifarelli, D. M. (2003). VaR as a risk measure for multiperiod static inventory models. *International Journal of Production Economics*, *81-82*, 375–384. doi:10.1016/S0925-5273(02)00369-9

Mostard, J., & Teunter, R. (2006). The newsboy problem with resalable returns: A single period model and case study. *European Journal of Operational Research*, *169*(1), 81–96. doi:10.1016/j.ejor.2004.04.048

Özler, A., Tan, B., & Karaesmen, F. (2008). Multi-product newsvendor problem with value-at-risk considerations. *International Journal of Production Economics*, *117*(2), 244–255. doi:10.1016/j.ijpe.2008.09.014

Petruzzi, N. C., & Dada, M. (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research*, *47*(2), 183–194. doi:10.1287/opre.47.2.183

Tapiero, C. S. (2005). Value at risk and inventory control. *European Journal of Operational Research*, *163*(3), 769–775. doi:10.1016/j.ejor.2003.05.005

Vaagen, H., & Wallace, S. W. (2008). Product variety arising from hedging in the fashion supply chains. *International Journal of Production Economics*, *114*(2), 431–455. doi:10.1016/j.ijpe.2007.11.013

Wu, J., Li, J., Wang, S., & Cheng, T. C. E. (2009). Mean-variance analysis of the newsvendor model with stockout cost. *Omega*, *37*(3), 724–730. doi:10.1016/j.omega.2008.02.005

KEY TERMS AND DEFINITIONS

Linear Demand: Is a widely used price-dependent demand model which is a linear function of the retail price.

Multiplicative Demand: Is a widely used price-dependent demand model under which demand is the product of market parameters and the retail price.

Newsvendor Problem: Is a classical single-period inventory model with stochastic demand.

Price-Dependent Demand Newsvendor Problem: Is an extension of the newsvendor problem with price-dependent demand model.

Value-at-Risk (VaR): It is a useful and popular objective in risk analysis.

This work was previously published in Fashion Supply Chain Management: Industry and Business Analysis, edited by Tsan-Ming Choi, pp. 100-109 , copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 46
Implementation of Rapid Manufacturing Systems in the Jewellery Industry in Brazil:
Some Experiences in Small and Medium-Sized Companies

Juan Carlos Campos Rúbio
Universidade Federal de Minas Gerais, Brasil

Eduardo Romeiro Filho
Universidade Federal de Minas Gerais, Brasil

ABSTRACT

This chapter presents the rapid prototyping and manufacturing concepts applied as means to reducing time between jewellery designs and manufacturing process. Different processes on jewellery modelling production are presented. Nowadays, the use of technologies as CAD/CAM - Computer Aided Design and Manufacturing in high production companies are very disseminated. However, the implementation of these resources at the design and manufacturing processes of jewels and fashion accessories, in small and medium size businesses, is still insipient. As reference, is presented the situation observed in small and medium companies located in Minas Gerais, Brazil.

INTRODUCTION

This chapter presents an analysis of the use of CAD/CAM systems (Computer Aided Design and Computer Aided Manufacturing) aiming to improve the connection between design and production in small and medium size jewellery companies located in the state of Minas Gerais, Brazil. In addition to that, the changes caused by this technology in the professional relationships are also discussed. The principal point is related to the influence of the use of CAD/CAM system on the designers' skill (competence) and design methods in an interactive design process and a Rapid Manufacturing (RM) system implementation.

DOI: 10.4018/978-1-4666-1945-6.ch046

In a global and highly competitive market, it is critical for the companies to develop distinctive products, joining aesthetic features with the effective control of the production processes, in order to ensure improved results. Within this scenario, the designer's activities are becoming more important for the jewellery industry; nevertheless, the design process is frequently regarded as an artistic activity dissociated to the industrial process. Improvements in jewel manufacturing process are extremely important, because the jewels and fashion accessories presents a high aggregate value. It is most important when raw material is abounding too. Brazil is a precious metal and stones products country, namely gold, silver, diamonds, emeralds, etc. Thus, better products imply in leadership of the market and consolidate designer's names in global market.

The intensive use of rapid manufacturing and CAD/CAM systems brings new opportunities to the integration between design and manufacturing process. Through a better assessment of the products during the preliminary phases of the project it is possible to predict mistakes that could result in additional labor, as well as increased production time and cost. In spite of technological advances and the use of machines at several stages of the industrial production of jewels (conception and execution of the first model), the process still maintains essential handcraft features. The intrinsic limitation of the manufacturing processes used in this phase for designer and goldsmith, results in mistakes which are transferred to the next phases, impairing process control and manufacturing phase. This is an important aspect to be considered in the product development, especially in small and medium size companies, which are typical of the jewellery industry in Brazil.

Changes in the communication methods and technical tools, such as CAD/CAM systems, must be carried out simultaneously with changes in the designers work method, aiming to adjust their projects to the needs of other production sections (including the goldsmith), with the purpose of obtained a finished product with the required quality. The RM concept can be used when the development of moulds or tools for moderate volume parts or products is required. Among the various RM processes, computer numerically controlled (CNC) machining is recognized as the technology most widely used in industry for the application of CAD/CAM systems. The production of jewels is based on the investment-casting process, in which the moulds and casting pattern can be produced in plastic or wax using an RM technology, thus reducing time and cost as well as ensuring the production of goods with high quality.

Different researches (as Rocha, 2007, Fernandes et al., 2005 and Han Boon et al., 1992) indicates that the use of Computer Integrated Manufacturing (CIM) systems can offer a substantial assistance to the solution of issues related to the interface between design and manufacture of jewels. However, some drawbacks associated mainly with the lack of qualified professionals (especially designers) were identified (Siu & Dilnot, 2001). The use of these technologies is only possible by employing three-dimensional CAD systems and trained professionals able to handle this task are currently not available in the jewellery industry. Moreover, the application of these tools alone is not enough to ensure the satisfactory integration of all phases of the project.

In jewels international market the Brazilian designer and manufacturer are traditionally competitive at producing small stones fashion jewellery and their skills in gem setting. These products attend the needs of its customers in terms of originality, creativity and fashion. However, the Brazilian Jewelers needing to search for ways to increase the efficiency and improves the productivity. Thus, this Chapter is concentrated on study of the rapid manufacturing innovative technique and their use in jewellery production for Brazilian small-size companies.

DESIGN AND MODELING PROCESS IN THE JEWELLERY COMPANIES

The first step of jewellery industry production process is the concept and design of a new jewel, normally by a specialized industrial designer. After design process conclusion, another professional, the goldsmith, creates a prototype that will be the base for all subsequent industrial process. The interaction difficulties observed between designers and goldsmiths in the design and production stages are various and bring many problems to production results.

The deficiencies observed in the information transference at initial stages of jewel design and production results in important consequences for next phases, including serious imperfections in the final product, re-work in the production process, increasing of operational costs and delays in delivery product. In the same way, the improvement in the initial phases of jewel production represents a global development of all process quality results. Special attention is necessary on the design product process, for the reason that in this phase are elaborated most important specifications to an adequate and balanced production process. The design decisions are fundamental to a better quality in the final product and the success in a highly competitive market.

The Design Process

The designer activity begins with the creation of a piece or collection jewel. Ordinarily, the project includes specifications about costs, market share, piece weight, materials etc. From this information, the first ideas and freehand sketching drawings are doing on paper. After that, an adequate solution selection between different proposals is selected and refined in another development phase, a detailed and formal design plan. This result is sent to production phase (Figure 1). Conventionally the jewel design includes a perspective view (in colour, with an "artistic presentation", when is possible to anticipate the formal result, close to the real appearance of jewel to manufacturing) and a technical drawing (with the essential view to execution, in most times, superior and lateral views) that includes measure and materials used, finishing, enlarged details etc.

However, this formal representation, including perspective and artistic representation, technical drawings and others technical information, sometimes is not completed. In majority, the designs are presented to goldsmith using partial or alternative draws (Figure 2), unstudied techniques in different forms:

- Colour draw with one side of the piece, accomplished with measure-simplified representation;
- Technical drawing, using ISO norms;
- Technical detail with material and dimension description;
- Sketches and drafts;
- Prototypes using alternative materials (clay, paper etc.)

About representation ways, Noguchi (2003) presents a designer interview: "Some pieces are illustrated using technical draw, surfaces, cuts etc. A clay model is used when the goldsmith is not a high skilled professional. Sometimes, when I am tired of drawing, I do only a mock-up, the "direct form". Other times I do only a "sketch" on paper. When the goldsmith is highly skilled, he almost knows the future result, if it will be satisfactory".

Figure 1. A final jewel design (including details)

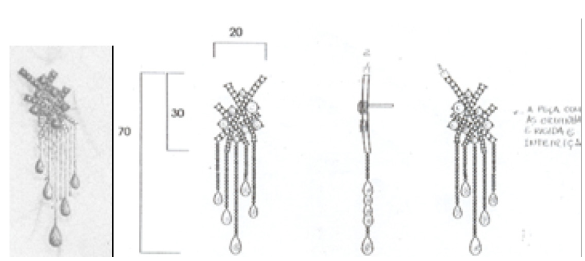
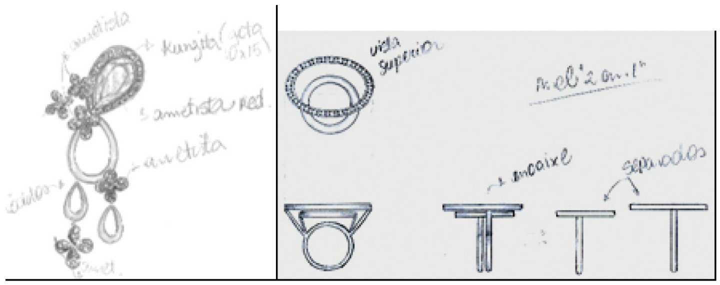


Figure 2. Jewel Design Examples. (Left: Design including only a colour draw, with material indication. Right: Design with only a 3 views draw, including fittings solutions. Source: Rocha, 2007.)



The Figure 2 presents some examples of "unconventional" designs, released by jewel designers.

The most important communication element between the conception and production steps is the jewel design. The quality and rules used in design process are important, however sometimes the information included in design documentation are not sufficient to goldsmith comprehension and don't trust the consistence in the process, requiring others representations (and communications) resources. Eventually, some design characteristics are "deduced" by goldsmith, because cannot be presented on paper as like as in a three-dimensional model. Movements, fitting systems, volume and weight of pieces are items not always adequately represented in the conventional draw. Moreover the characteristic limits on the representation process, there is the problem of the ignorance by designers about the jewel production process, its possibilities and restrictions. This fact can cause a series of incompatibilities between design and production, bringing delays and changing necessities in the pieces.

For this reason, it is usual that modelling goldsmith modify some design characteristics, adapting it to product process limits and avoiding problems in the final model. These modifications can respect and maintain the original design aspects, when are about specific adaptation as plate thickness, measurement or fitting adjust etc. However, in some cases the goldsmith change important aspects of design pieces, owing to impracticability in the production process or about difficulties in the design interpretation.

The Wax Model

After conclusion of design work, the draws are sending to goldsmith, which start the construction of a "model" (or prototype) of jewel piece, in plastic or special wax. This model can be made by metal (gold, silver etc.), plastic or a special wax (used in jewel industry). The choice depends fundamentally of the techniques used in the piece confection. In this process are used specific tools like files, saws, drills etc. in a handcraft process, until the final form of the piece.

In the traditional process, the model construction is a fundamental step at industrial process. In cases of exclusive pieces, made to order, the goldsmith finalizes his work producing the first piece directly in metal. But in the industrial process, thousands pieces can be produced from the wax model. This step (model construction) is a handcraft work and depends on highly of the goldsmith skill. For this reason, problems can be caused by limitations in the work conditions and tools. Some design pieces propose details and symmetries impossible to handcraft techniques. In the same way, complex geometric parameters turn impossible an adequate control by goldsmith about some aspects like exact weight and thickness in all points of piece, in order that the final product will be different that described in the design.

REVIEW ON JEWELL RAPID PROTOTYPING

In recent years, the use of computers throughout all engineering disciplines has obtained an increase importance. This is particularly true in activities that span the life cycle of discrete products development, namely in products design and in manufacturing. Computer-aided manufacturing design (CAD) and computer-aided manufacturing (CAM) software programs, computer numerical control (CNC) machining and Rapid Prototyping (RP) process has certainly gained momentum in the marketplace.

Several works have been published aiming to evaluate the RP process and computer technologies on the manufacturing mechanical and medical complex pieces. However, a few studies have been making of the RP techniques and CAD/CAM/CNC tools in jewellery industry. Leong et al (1998) showed the effect of abrasive jet deburring processes on the surface finishing of jewellery models built by stereolithography apparatus (SLA). The investigation aims to determine the significant parameters of the deburring process, to determine a practicable range of settings of these parameters for effective deburring, and to establish the optimum settings to the jewellery models built by stereolithography.

Traditionally, a human process planner performs the manufacturing process planning, with acquired expertise in machining practices that determines from a parts engineering drawing what the machining requirements are. Wong et al (1999) developed a decision support system to assist in the operations of typical jewellery manufacturer. The paper describes the development of a management information system with a specially designed decision support system to meet the specific requirements of Hong Kong's jewellery industry. It aims at assisting a jewellery manufacturer to make decisions in various areas of operation, including price quotation, sales analysis, materials requirement planning, vendor selection and job assignment.

Using the conventional design and production system in the Hong Kong jewellery industry as a case, Siu and Dilnot (2001) investigated the reason for the high degree of failure of design representations in the jewellery context, through illustrating the difficulties and problems (in maintaining quality of craftsmanship) when introducing CAD/CAM/RP technologies in production of craft-based object as jewellery in modern factory production. In this study, the authors affirm that the tacit-format attributes, which were separately contributed by the jewellery designers and goldsmith, can be extracted, recaptured, recorded, integrated and finally coded into CAD database.

Many commercial CAD systems have been developed for the purpose of designing jewellery (e.g., JewelCAD, JewelSpace, Rhinoceros, Matrix, ArtCAM Jewelsmith, etc). They provide graphical interfaces with excellent rendering capabilities. The majority of these systems provide built-in libraries of settings and cut gems and stones and advanced feature-based design tools. Also, the majority of these systems have the capability of exporting models to RP Machines. However, none of these systems is appropriate for designing and creating pierced Byzantine jewellery or other jewellery of this kind of detailed engraving. Stamati and Fudos (2005) shows a parametric CAD system for the design of pierced medieval jewellery, which is jewellery created by piercing, a traditional Byzantine technique. For other hand, Soo et al. (2006) applied fractal geometry for artistic product design in a CAD platform. Based on the proposed data structure, the geometrical information of the fractal can be extracted in a more effective way and the fabrication of a fractal shape for an artistic product design can be realized. The CAD platform has been employed to fabricate a physical prototype in RP technology.

Recently, in the paper titled "I want and now", Langer (2008) affirm that consumers demand for individualized products. This individualized series production, or call it mass customizations, is driving a paradigm shift for manufacturing and electronic manufacturing (e-manufacturing) with its flexible and fast response times is a major factor in the drive for change. The e-manufacturing allow delivers end products, functional parts and tools directly from computer aided design (CAD) data. A laser heats and melts powdered metals or plastics layer by layer, until the build is complete. Among the novel products with highly complex and sometimes filigreed structures and forms that are unthinkable geometries for conventional series production is jewellery, clothes, lamps, chairs or others functional parts for industry.

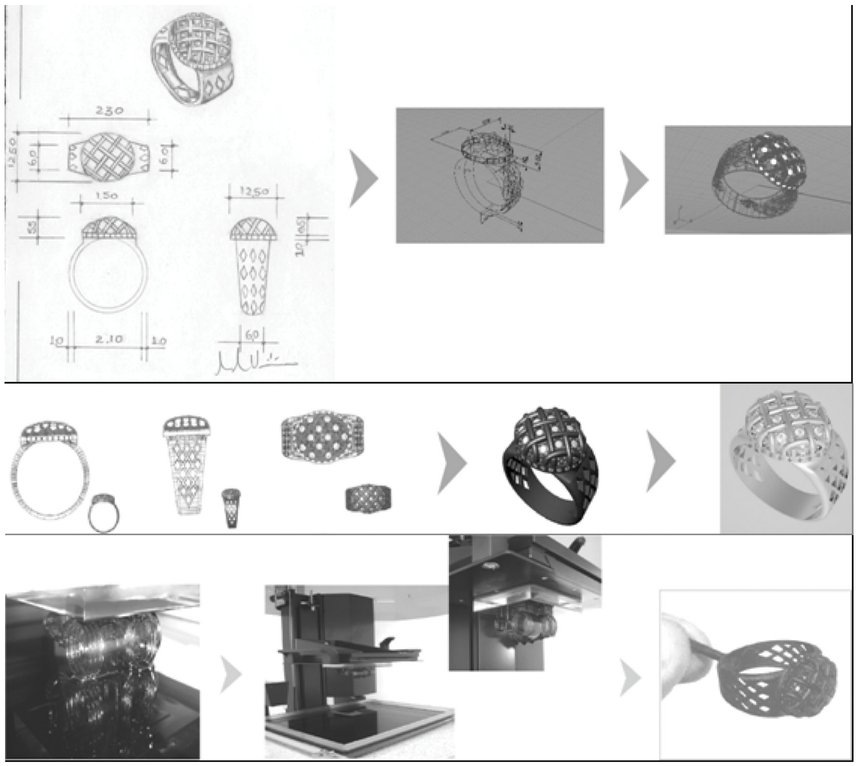
NEW TRENDS ON CAD/ CAM IMPLEMENTATION IN JEWERLY INDUSTRY

As far as concerned on computer technologies, the resources are today very popular and used in different industrial sectors since 1960 decade. At first in the aerospace and automotive industries, the CAD/CAM (Computer Aided Design and Manufacturing) has been expanded the application on jewellery. CAD Software, CNC (Computer Numeric Control) machines and Rapid Prototyping equipments are improving the jewellery sector results, bringing new competitive advantages. On the other hand, some questions obstruct an accelerated diffusion of new technologies between the users, especially in the medium and small companies. These questions are highly related to the difficult in the learning process by designers and the conflict between using the new CAD/ CAM technologies and the traditional work process, including the relationship between designers and goldsmiths.

The improvements on results of new technologies application are only possible, in practice, with an intensive use of Three-dimensional CAD Systems (CAD3D). However, the professionals with a high knowledge about CAD are almost rare, and can be observed a resistance by traditional designers in front of new tools, in general more complex and (initially) slower than the traditional sketch process. The sub-utilization and unfamiliarity with the real potential of CAD are, as a rule, the main reason of the problems observed. In fact, many designers put questions about use of CAD technologies instead of traditional sketches and manual draws. The CAD application needs a high detailed solution, which can need more time in the initial steps of design process. However, the reduction of mistakes in the next steps and the improvement of a global result, owing to a better evaluation of design alternatives, can result in a time and resource economy in a long time. These aspects justify an effort in learning and application of these systems in the product development and production process. Notwithstanding the apparent complexity, after the software learned, the total time necessary to design process can be significantly reduced with facilities to generation of new alternatives, for example.

In the midst of new exigencies of consumers, quality improvement, reduction of products life cycle (time in market) and technical innovations, the jewellery industry has slowly introducing resources as CAD/CAM. By means of its characteristics, these systems present potential benefits in a crucial problem in jewel design: the communication interface between design and production sectors. The use of CAD system associated to CAM tools and rapid prototyping technologies turn possible a total integration in the initial steps in jewel production. A CAD3D system using solid modelling is able to simulate a virtual (and three-dimensional) model, showing forms and dimensions exactly as a real piece, including elements as colour, stones and construction details. These characteristics can be anticipated and visualized in the conception stage, avoiding problems like inaccuracy interpretation, improving quality in final design and reducing fails in the production process (Figure 3).

Figure 3. An example about a jewel development process: (a) handmade sketches; (b) first studies using a CAD system; (c) wire frame perspective; (d) different views in CAD; (e) jewel perspective; (f) jewel materials simulation in CAD3D; (g) and (h) model production;(i) final model in special resin.



Using a CAD3D virtual model, the designer can visualize the final appearance of jewel, with a “rendering” resource. It is possible to simulate a change of the material (for example, testing the results in same piece applying gold or silver as material basis). The modern CAD technologies present simulations very close to the real product. Different evaluation versions can be generated quickly, without wax models or different draws. The CAD systems bring great benefits in highly detailed pieces, when the exact dimensions and mechanisms descriptions are essential. Fitting and fixation mechanisms in these models are, in practice, impossible to a goldsmith. Also in the repetition of components in different pieces (as a collection), is possible by achievement in computer memory. At long last, CAD system can automatically generate all technical information essential to production and directly send to CAM machines, evicting mistakes interpretation.

After design process conclusion by CAD, nevertheless is possible a traditional production process, with the goldsmith work. In this case, the model can be printed, in different views, including technical and “artistic” information. However, the CAD resources are better used in connection with CAM technology, especially Rapid Prototyping machines to models construction. The main advantage of this technology is the possibility of a controlled result; with sure of the model resultant have the same characteristics of designer conception. Another fundamental advantage is a possibility of adoption of complex and symmetric forms, impossible to a traditional process production, involving a goldsmith work. Finally, it is observed a sensible reduction in the time-to-market, since the design and production process are integrated and the time between conception and production is progressively reduced.

METHODS USED IN JEWELLERY MANUFACTURING

Several techniques have been employed to make jewellery from gold and silver and its alloys are Investment Casting, Die casting, cuttle fish casting and hand fabrication. Many jewelers use CAD/CAM to make jewellery. CAD (Computer Aided Designs) and CAM (Computer Aided Manufacturing) to reproduce a model piece of jewellery that can be mass-produced. Making jewellery requires knowledge and expertise in gold smiting, stonecutting, engraving, mold making, fabrication, wax carving, lost wax casting, electroplating, forging, and polishing. These are the various steps needed to make jewellery. The first step in making a detailed piece is making of a mold (Ray, 2009), it can be made in desk or computer aided.

Mold

A mold is the exact and perfect replica of the piece to be made, copied either from a design or a piece/object. A mold is shaped around the shape/figure with the help of casting process. Molds are made from many materials, including plaster compounds. Some different methods of casting include the lost wax process, centrifugal or investment casting, and sand casting. In jewel manufacturing the process begins with a mold. The casting process involves a number of steps. There are two methods of casting using in jewel and fashion accessory manufacturing; investment casting or die casting each with its own advantages.

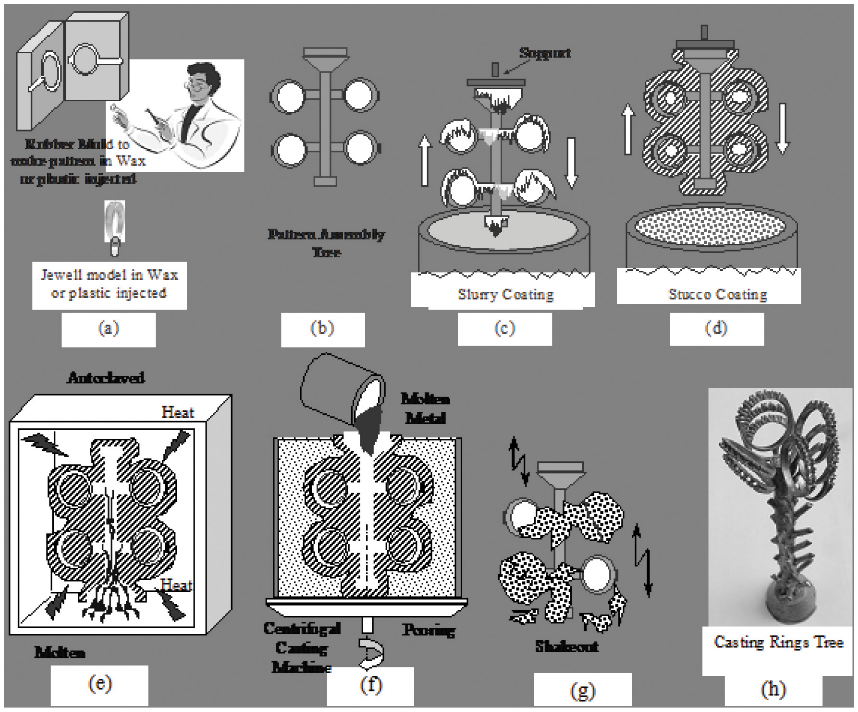
The Investment Casting

The "Investment Casting" or "Lost-wax Casting", also called by the French name of *cire perdue* or the Latin *cera perduta*, is the process by which a bronze, brass or other special alloys are cast from an artist's sculpture and jewels. The casting process is one of the oldest known metal-forming techniques, was first used during the period 4000-3000 B.C. Lost-wax casting was widespread in Europe until 18th century, when a piece-mold process came to predominate. Believed to have been developed by the Mesopotamians, it remains as the most popular process of making gold (metal) jewellery and forms the basis of modern investment casting process. Nowadays, with the appearance of the high-technology waxes, refractory materials and specialist alloys, the castings allow the production of components with accuracy, repeatability, versatility and integrity in a variety of metals and high-performance alloys, thus, in industrial uses, this modern process is called investment casting.

Kalpakjian (1995) said that this process involves dipping a mold into a ceramic mix. Sometimes new materials like plastic or polystyrene foam is used instead of wax. This process has a number of steps involved. The Figure 4 shows a brief description of the Investment Casting process begins by the injection of wax in a metallic mould that has previously built up or plastic hand-made pattern.

a. A primary model can be made in hard alloy like nickel silver or just silver. Surrounding this primary model, using sheet rubber in a mold frame, makes a rubber model. Placing it in a heated press then vulcanizes it. On cooling, it is cut with a scalpel into halves or more and removing the primary model. This rubber mold is used to make many copies of the primary model on wax. Molten wax is then introduced into the mold cavity by using a wax injector. On cooling the wax is removed to get an exact copy of the primary model in wax. Other way is wax or plastic models hand-made by craftsmen.

Figure 4. Steps involved in jewellery manufacturing by investment xasting process



b. Then the wax or plastic models are welded to a structure also in wax, to set up an assembly (tree form).

c. The ceramic mould or negative of the wax pattern is formed up by immersing the assembly (tree) in a slurry coating (babbitt), keeping the finest details as the wax pattern.

d. The assembly is then surrounded by refractory materials (e.g., fine silica) till that a ceramic shell is obtained, also called stucco coating or investment plaster. Thus, form stiff and sturdy molds are produced.

e. The wax is removed in an autoclave by a pressure-temperature process to pattern melt out. The flask is inverted and placed in a kiln/furnace. The wax is melted by steam or air to remove all the wax. The furnace is set in stages and the maximum temperature reached is 750 degrees centigrade. The melting process takes about 12 to 16 hours. This melting down of the wax is called the 'lost wax process'.

f. The ceramic mould is pre-heated at 1000°C and the metallic material is poured into at temperature higher than fusion temperature (e. g., 1550 °C steel or 1100 °C copper, etc) Two types of casting machines can be used; the centrifugal casting machine which is the older technique or the modern technique of static vacuum assist machines.

g. When the assembly is cold the ceramic shell is separated from the steel bye vibration process.

h. The steel parts are then separated for the assembly by mechanical process.

i. The process finishes with all kinds of controls, heat treatments and machining operations to achieve quality and finishing required.

As advantages of casting investment in jewel manufacturing can be said that it is an old age proven method. It allows the jewellery flexibility to create complex designs. The details can be copied perfectly. The control of colour is better. The finished product can be highly polished. It results in very fine surface finish. The metallurgical properties are also excellent. However, some disadvantages of investment casting also can be cited, as that this process can result in porosity. Also the dimensions may not be as accurate as the die struck method. Nevertheless, this process can and is used for almost all gold jewellery and remains a favourite with jewellers even after 6,000 years later.

Die Struck Method

Die struck method is a casting method where the metal to be cast is forced under pressure into a mold which is usually made out of metal. This is an excellent method of producing complex shapes. The earliest recorded history of die casting by pressure occurred in 1800's. Using a plunger or compressed air, molten metal is forced into a metallic die and the pressure is maintained until the metal settles and solidifies. The pressure reaches 300 MPa. The intense pressure causes the atoms in the metal to move closer together and solidify to dies form or molds. Using compressed sheet metal and dies steel mountings are formed with metal parts mechanically stamped out. Each part is matched and fitted into the correct portion of die halved and stamped and shaped. A hydraulic press is used (Ray, 2009).

MODERN MANUFACTURING SYSTEMS

A number of steps are involved in manufacturing a part from its conceptualization to production. They include product design, process planning, production system design, and process control. Computers are used extensively in all these stages to make the entire process easier and faster (Mehrabi et al, 2002). Potential benefits of using computers in jewel manufacturing can include reduced costs and lead times in all engineering design stages, improved quality and accuracy, minimization of errors and their duplication, more efficient analysis tool, and accurate control and monitoring of the machines/processes, etc.

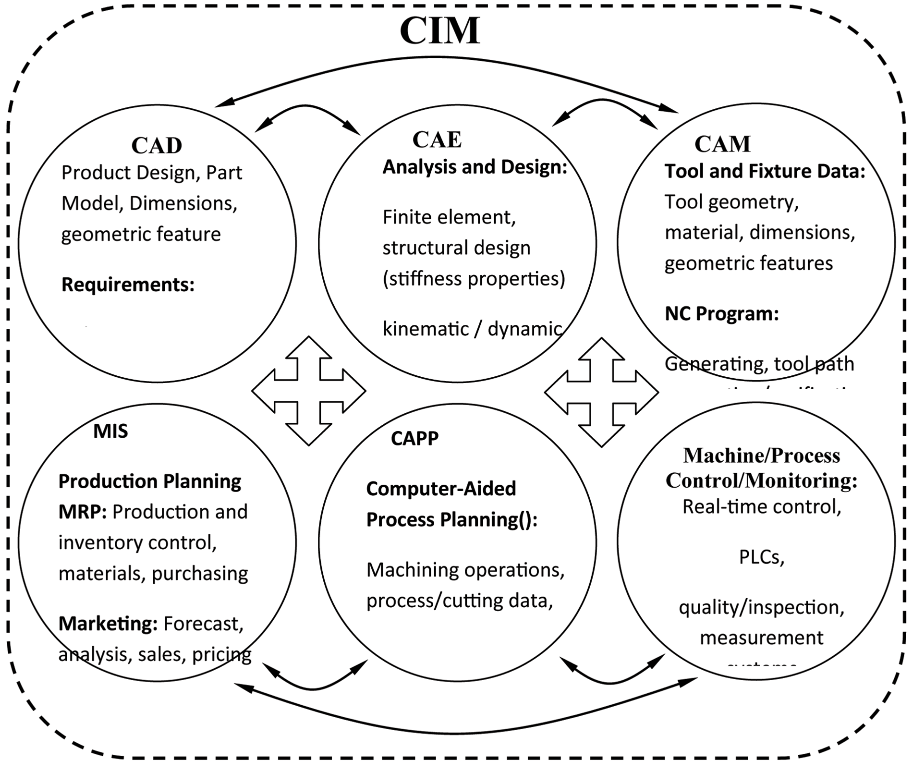
As showed in Figure 5, the applications of computers in manufacturing systems allow development products by computer-aided design (CAD), using computers in the design and analysis of the products and processes. The computer technologies play a critical role in reducing lead-time and cost at the design stages of the products/process. Also, computers may be utilized to plan, manage, and control the operations of a manufacturing system: computer-aided manufacturing (CAM). In CAM, computers are used directly to control and monitor the machines/processes (in real-time) or used off-line to support manufacturing operations such as computer-aided process planning (CAPP) or planning of required materials. Materials Requirement Planning (MRP) determines a schedule for the operations and raw material purchases shows the production requirements for the whole production system.

At higher levels, computers are utilized in support of management. They play a critical role in all stages of decision-making and control of financial operations by processing and analyzing data and reporting the results (Management Information Systems, MIS). Also, computers facilitate integration of CAD, CAM, and MIS (Computer-Integrated Manufacturing, CIM). They provide an effective communication interface among engineers, design, management, production workers, and project groups to improve efficiency and productivity of the entire system. (Mehrabi et al, 2002).

Product Development Life Cycle

Many enterprises have new product development methodologies that provide a standard framework for planning and managing development efforts. It can be discriminated four phases to this life

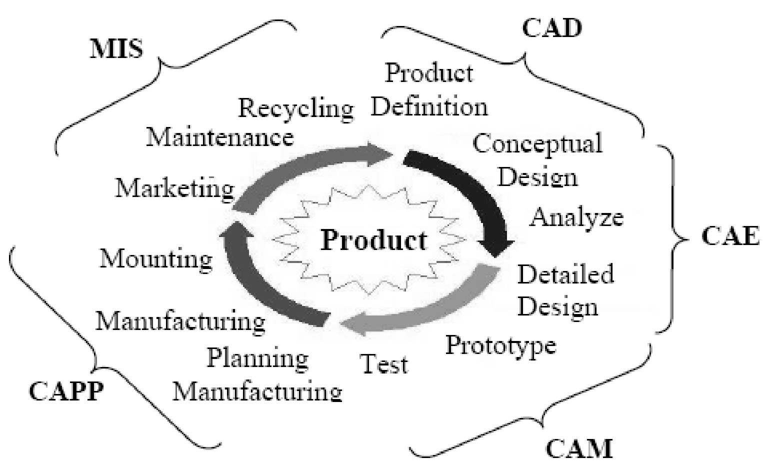
Figure 5. Applications of computer technology in manufacturing



cycle: concept (preliminary ideas), production (including product development and purchasing), use and disposal. The Figure 6 shows the life cycle phases and computer tools that can be using in improvements of the process.

Improvements in jewel manufacturing process are extremely important. Because the jewels and fashion accessories presents a high aggregate value. It is more important when raw material is abounding too. Brazil is a precious metal and

Figure 6. Products life cycle phases and computer technology applications



stones products country, namely gold, silver, diamonds, emeralds, etc. Thus, better products imply in laddering of the market and consolidate designer's names in global market.

An Integrated Method for Jewel Manufacturing

For the last few years, the process of manufacturing jewellery has seen lots of innovation, and particularly in term of Design & Manufacture. CNC machine are adapting to the jewellery industry, as for the laser welder and market, and Design & Manufacture software. For the last one, it has come much slower, and the available solutions offered on the market are not affordable to the majorities of the Brazilian jewelers.

Rapid Prototyping, RP, techniques has a vast application on several fields, such as architecture, art, engineering and industrial design. Using these technologies, and from a CAD3D geometrical file, it is possible in a few hours to build functional prototypes of components or even production tools, known as Rapid Tools or Rapid Manufacturing (RM). Hence, there is no need to build moulds to produce some components, e.g. in the development stage, that are constantly changed by the design teams. The number of RP systems in use, as well as theirs typology, is increasing every day, using different building materials such as waxes, polymers, metals, paper, sand and mix materials. There are already some desktop solutions (Mascarenhas & Esperto, 2006). Using two (2) of these RP techniques it was created a few jewel prototypes, namely Fused Deposition Modeling (FDM) and CNC micro-milling machining.

Rapid Manufacturing is the embodiment of the concept that allows defining elements or modules with well defined interfaces and "hooks" in a way that the process of manufacturing can be: developed in an independent manner, customized based on installation specific data, based on formal models of controller behavior, amenable to automatic generation of control software modules, and reconfigurable to adapt to changes in the environment through changes to data and/or regeneration of control software. Smith (1998) shows that application of RM allows identify time saving opportunities in the product development lifecycle, namely in design and manufacturing phases. As showed in Figure 7, the jewel manufactures may be improved by reducing both the production time and costs by work's superimposing. It is possible by using the computer technology, namely Rapid Manufacturing (RM).

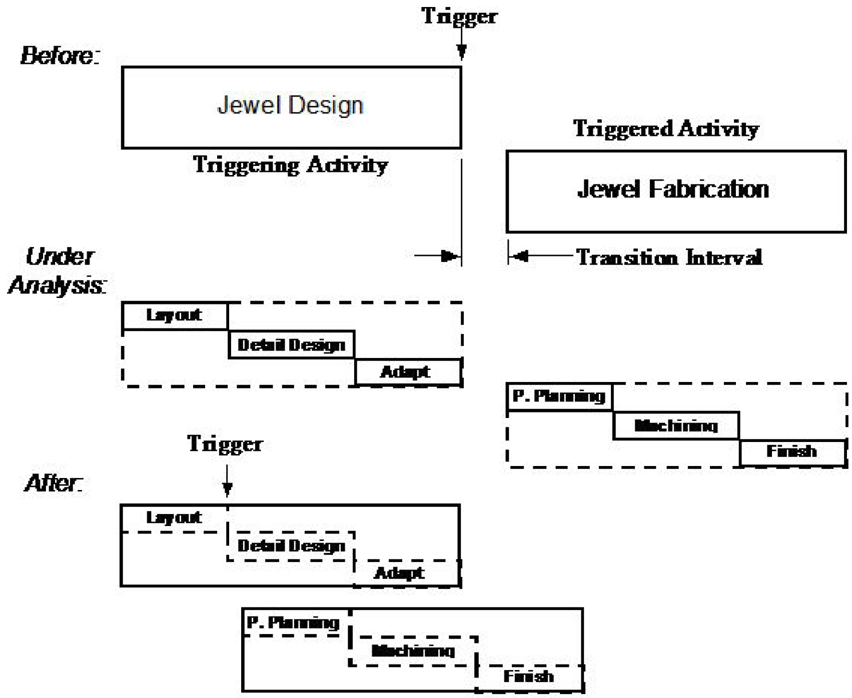
In last year's, artistic design & manufacture software has been developed to the jewellery industry, aimed to reduce production costs and hand process. This technology allows designers to the freedom to development quickly design, modify and view the models. Thus, designers can create new models with the benefit of a parametric construction tree which allows modifying the design at any stage, without going back countless steps or starting from scratch, reducing both the production time and costs. It can use the jewellery specific tools, such as the ring wizard to quickly build a band then change for example, the finger size while keeping the thickness constant.

Once the jewel model is completed, the designer can resize it, reconfigure its dimensions or use it as a component in another design by adding it to a pre-existing library of parts. Another problem is the cost of the jewel designed; the software can alert the designer if it goes under or over the desired weight. Allow a more accurate vision with the production costs. Most of materials are in the software, and designers can also easily create your own material for new jewel or fashion accessories.

Methodology Using in Jewel Rapid Manufacturing

The Rapid Manufacturing (RM) concept can be used when the development of moulds or tools for moderate volume parts or products is required. Among the various RM processes, computer numerically controlled (CNC) machining is rec-

Figure 7. Time saving opportunities in the product development life cycle



ognized as the technology most widely used in industry for the application of CAD/CAM systems. The basic methodology for RM techniques can be summarized as follows:

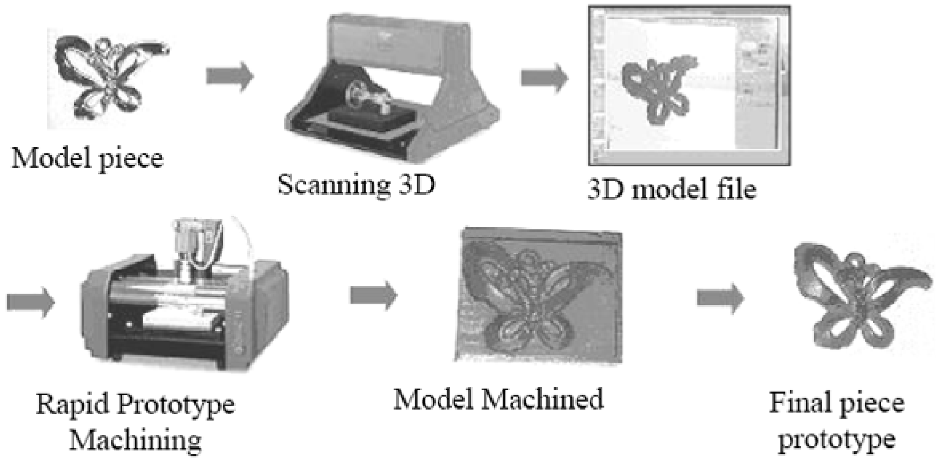
a. CAD jewel model is constructed, and then converted to STL format. The resolution can be set to minimize stair stepping.
b. The RP machine processes the STL file by creating sliced layers of the model.
c. The first layer of the physical model is created. The model is then lowered by the thickness of the next layer, and the process is repeated until completion of the model.
d. The model and any supports are removed. The surface of the model is then finished and cleaned.

Different CNC machines can be used in rapid prototyping aided jewel manufacturing. In Figure 8, a small Roland desktop 3 axis machine was used as RM process, It can machining soft metal, wax or resin that can be vulcanized in rubber moulds. There are two work ways; from the scratch (start drawing) and 3D scanning and modelling (reengineering).

CAD/CAM IN JEWELL INDUSTRY

The use of latest trends in design tools and manufacturing technologies allows jewels design and manufacturing through CAD/CAM technology, from the basic concept to final productions. According to Osgrove (2005), can be cited three (3) advantages of this technology in jewel industry, namely accuracy, time and cost advantages. Among of the accuracy advantages that can be cited are the capability in provides a realistic image of the jewel even before it is made, thus aiding inspection to improve quality. Same, achieve exact symmetry of design, especially for left and right

Figure 8. Diagram illustrating the various methodologies' phases used in jewel rapid manufacturing



earrings. It can create identical looking designs of different diamond weights fasters. Likely, can be created ensembles will have same family looks, described precisely the metal dimensions for gemstone holding areas, no variation in consistency and accuracy, such as provides casting friendly even cross sections, etc.

Regarding to the time advantages in the application of the CAD/CAM technology in jewel manufacturing, can be said that CAD and CAM based technologies help to obtain a silver master models, faster than traditional wax or solid silver based carving methods. One can convert a complex design sketch into a finished silver model in few days time. Modifications in master model are faster and creating identical designs for different diamond weights, earrings and ensembles is possible.

The costs are extremely important in jewel manufacturing, and CAD/CAM technology allows obtaining the uniformity of scoop for maximum removal of excess metal, giving consistent and minimum metal weight. Same, the model making productivity increases many times without increase skilled manpower. As form a contract of high capital involvement, optimum utility of the capacity makes models affordable is allowed when CAD/CAM technology is applied in jewel industry. By using the CAD software in jewel development the designer might creates and produces a bigger range of jewellery models at the same time (family pieces) allowing creating and producing many finished models simultaneously.

In practice, CAD software is written in numerous different programming languages and can be used in many environments and platform (e.g., Windows, Linux, Apple Mac, etc). Some CAD using in jewellery are written in the Java computer, they can also run in any machine, and exchange data with other users. Jewellery's CAD software present friendly and intuitive interface, as well as help box dynamic. There are libraries of precious metals and gemstones with gemmological materials, which can also add too. Thus, a designer with few days of practicing would be able to design nice items, showing that working on the model can be very clear and the final rendering to be easy to use and extremely effective.

Kai et al (2000) presents a methodology to converting scanned data to.STL format. They said that the data modeller is the most important process for the success of jewel making. Its purpose is to convert the scanned image file to editable CAD format. The common converted file formats include IGES, DXF and ACIS. The conversion to CAD readable format is to facilitate correction and rework of the scanned image before converting to.STL format. Figure 9 illustrates the flow chart for converting scanned data to.STL format, followed by Figures 10a and 10b which present the steps to convert the input (scanned file) to the desired output (.STL file) of the ear mould, and finally can be used as a wax jewel model or vacuum cast mould.

All RP systems generally adopt the additive fabrication technique, which is a process that builds a part layer by layer. Each layer is represented and determined by its corresponding slice file. The slice file stores the geometric detail data of the model at that particular slice height. The slice file is created from the.STL file. After specifying the slice height, usually smaller 0.30mm, the RP attached software will slice the. STL file to a slice (.SLI) format. The.SLI file is then used to build the RP model (Osgrove, 2005).

JEWEL MODEL RAPID PROTOTYPING

Rapid Prototyping (RP) can be defined as a group of techniques used to quickly fabricate a scale model of a part or assembly using three-dimensional computer aided design (CAD) data. Rapid Prototyping has also been referred to as solid free-form manufacturing; computer

Figure 9. Steps indicating the conversion of scanned data to.STL format

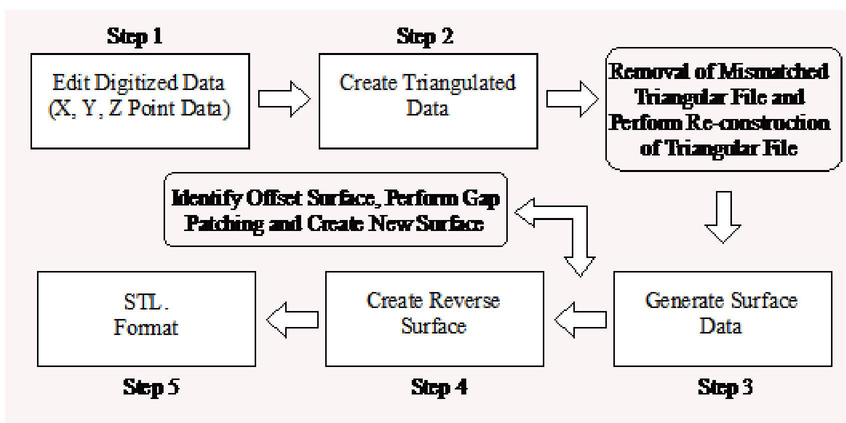
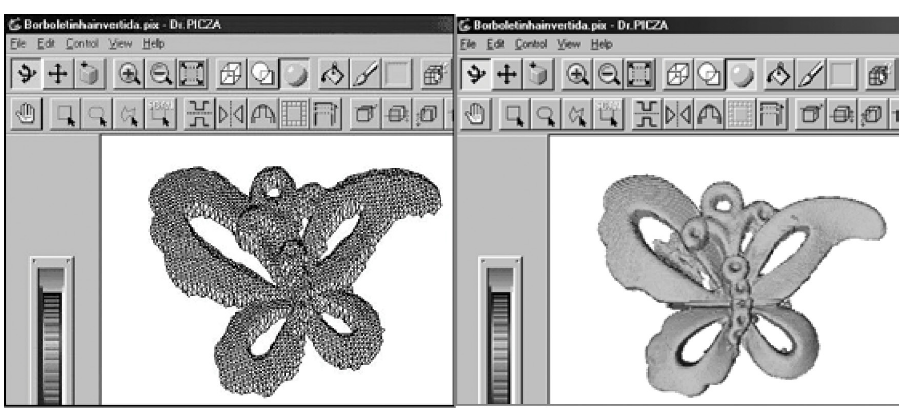


Figure 10. Convert; (a) the scanned image file to (b) editable CAD format



automated manufacturing (CAM) and layered manufacturing. RP has obvious use as a vehicle for visualization. In addition, RP models can be used for testing, such as when an airfoil shape is put into a wind tunnel. RP models can be used to create male models for tooling, such as silicone rubber molds and investment casts. In some cases, the RP part can be the final part, but typically the RP material is not strong or accurate enough. When the RP material is suitable, highly convoluted shapes (including parts nested within parts) can be produced because of the nature of RP.

Why Rapid Prototyping?

The reasons of applied Rapid Prototyping (RP) in jewel manufacturing are:

- Increase effective communication.
- Decrease development time.
- Decrease costly mistakes.
- Minimize sustaining engineering changes.
- Extend product lifetime by adding necessary features and eliminating redundant features early in the design.

Rapid Prototyping decreases development time by allowing corrections to a product to be made early in the process. By giving engineering, manufacturing, marketing, and purchasing a look at the product early in the design process, mistakes can be corrected and changes can be made while they are still inexpensive. The trends in manufacturing industries continue to emphasize the following:

- Increasing number of variants of products.
- Increasing product complexity.
- Decreasing product lifetime before obsolescence.
- Decreasing delivery time.

Rapid Prototyping improves product development by enabling better communication in a concurrent engineering environment.

Different CNC machines can be used in rapid prototyping aided jewel manufacturing. In this work, two RP systems were used to fabricate the jewel and fashion accessories pattern for the silicon mould making or wax model. They are fused deposition modelling (FDM) and CNC Micro-Milling Machining. The first works by material deposition and the second works by material retreat. The two RP systems will be briefly discussed in the following. In addition, SLA and CNC will be used to make the RP jewel mould or model. This is to facilitate the evaluation on the patterns fabricated by each RP system. The evaluation will show which RP systems produce better-finished pattern suitable for investment casting. Figure 11 shows development process of the ring using software oriented to RP machine inside RM environment.

Fused Deposition Modeling (FDM)

The Fused Deposition Modeling (FDM) process was developed by Scott Crump in 1988. The fundamental process involves heating a filament of thermoplastic polymer and squeezing it out like toothpaste from a tube to form the RP layers. The machines range from fast concept modelers to slower, high-precision machines. The materials include polyester, ABS, elastomers, and investment casting wax (Osgrove, 2005). The overall arrangement is illustrated in Figure 12.

Highlights of Fused Deposition Modeling:

- Standard engineering thermoplastics, such as ABS or investment casting wax, can be used to produce structurally functional models.
- Two build materials can be used, and latticework interiors are an option.
- Filament of heated thermoplastic polymer is squeezed out like toothpaste from a tube.
- Thermoplastic is cooled rapidly since the platform is maintained at a lower temperature.

Figure 11. Jewel development process inside RM environment

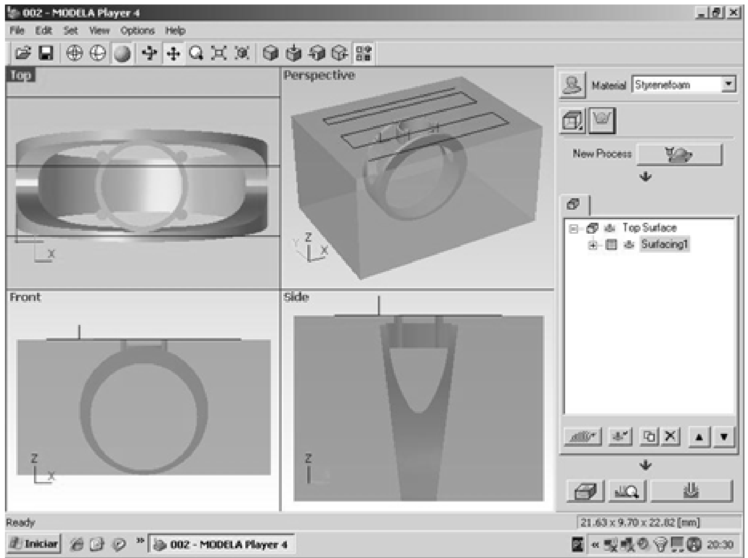
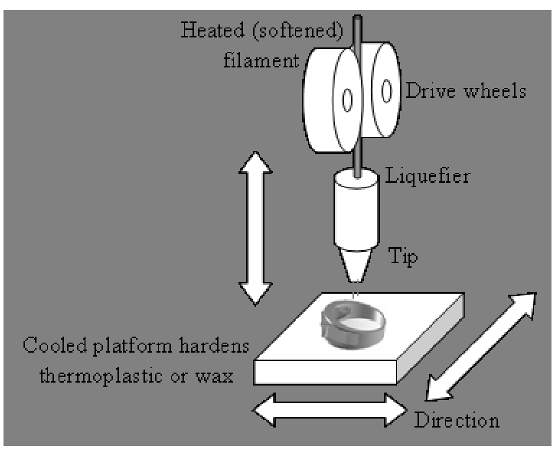


Figure 12. FDM extrusion head

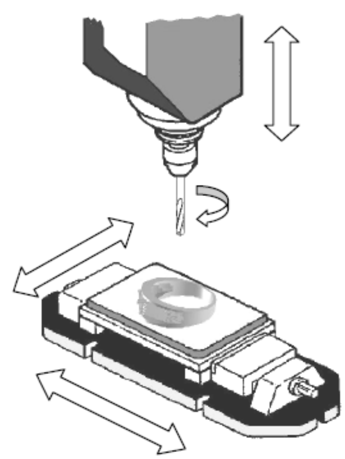


- Milling step not included and layer deposition is sometimes non-uniform so "plane" can become skewed.
- Make rapid progress in past few years and be used widely

CNC Micro-Milling Process

A wax model of a jewel can be produced using CNC machining techniques. The virtual model (CAD) of the jewel created from digital data obtained from the jewel model. The jewel model prototype is then fabricated from the wax model or other plastic materials. On the other hand, the CNC micro-milling process can be used in mold manufacturing. Thus, the virtual model (CAD) is used by coping the replica of the negative piece to be made, The Figure 13 shows CNC micro-milling machine used in jewel model rapid prototyping.

Figure 13. CNC micro-milling machine used in jewel rapid prototyping



Different CNC machines can be used in rapid prototyping aided jewel manufacturing. In this work, was chosen a small Roland desktop 3 axis machine, It can machine soft metal, wax or resin that can be vulcanized in rubber moulds. There are two works ways; from the scratch (start drawing) and 3D scanning and modelling (reengineering). Figure 14 illustrates the stages from scanned jewel model to wax jewel model and finally jewel prototype.

This alternative offers quickly turnaround metal mold prototypes and rapid prototyping for metal parts. Depending on material, timing requirements, critical dimensions and post-machining operations, there are several different methods that can be used to satisfy metal prototyping requirements. CNC machining is the most common method to manufacture rapid metal prototypes. Figure 15 shows a 3D model file and CNC machine cutting tool path using CAD/CAM software.

CONCLUSION

Everything considered, the benefits of CAD/CAM systems in front of the traditional design and production process are clear, and the use of these systems probably will increase in the next years. Notwithstanding the jewel industry modernization is only at the beginning. More and more references about this subject are presented, bringing incentive to entrepreneurs to implement new technologies in the jewel industry. However, the CAD/CAM is not a simple new tool, but it includes some difficulties to learn an adequate application. Problems like communication between designer and production sector, including goldsmith, can be solved with the use of CAD3D. In the same way, the technical characteristics of jewel are necessarily defined in the first stage of design process. This integration between design and production processes is the main benefit of the new technology. On the other hand, it is a great challenge to jewellery industry, especially small and medium enterprises.

Figure 14. Wax jewel model manufactured through CNC micro-milling machine

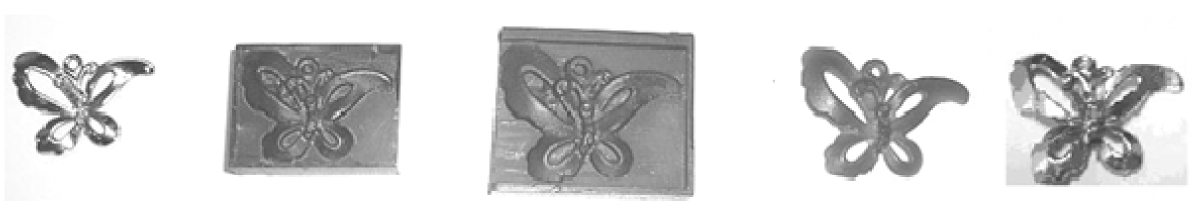
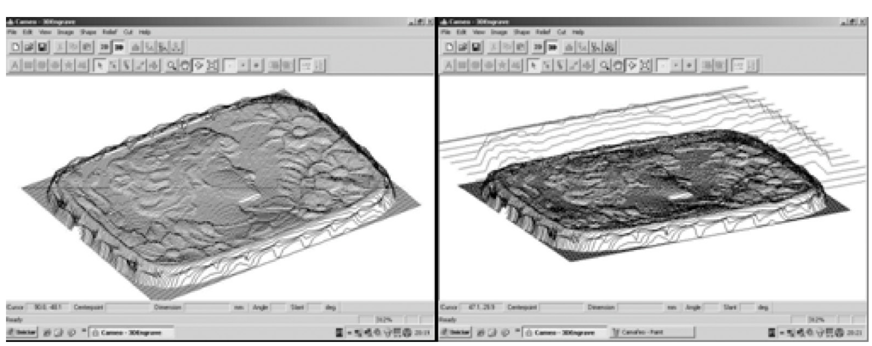


Figure 15. 3D model file and CNC machine cutting tool path using CAD/CAM software



REFERENCES

Fernandes, J. M., Romeiro Filho, E., Araki, L. A., Reis, L. P., Figueiredo, A. C., & Vaz, C. R. (2005). Cross-functional Teams and Concurrent Engineering: Contributions to the Development of Product Design Through Multidisciplinary Integration Using CAD Systems. *Product: Management and Development*, *3*(1), 5–9.

Han Boon, L., Ko, M. S. H., Gay, R. K. L., Kah, F., Leong, K., & Chua, C. (1992). Using computer-based tools and technology to improve jewellery design and manufacturing. *International Journal of Computer Applications in Technology*, *5*(1), 72–80.

Hilton, P. D., & Jacobs, P. F. (2000). *Rapid Tooling: Technologies and Industrial Applications*. New York: Marcel Dekker, Inc. doi:10.1201/9780203908020

Kalpakjian, S. (1995). *Manufacturing Engineering Technology*. Reading, MA: Addison-Wesley Publishing Company.

Kei, C. C., Meng, C. S., Ching, L. S., Tei, L. S., & Aung, S. C. (2000). Facial prosthetic model fabrication using rapid prototyping tools. *Integrated Manufacturing Systems*, *11*(1), 42–53. doi:10.1108/09576060010303668

Langer, H. J. (2008). I want it now - pushes manufacturers into rapid responses. *Metal Powder Report*, *63*(4), 8–9. doi:10.1016/S0026-0657(09)70035-9

Leong, K. F., Chua, C. K., Chua, G. S., & Tan, C. H. (1998). Abrasive jet deburring of jewellery models built by stereolithography apparatus (SLA). *Journal of Materials Processing Technology*, *83*, 36–47. doi:10.1016/S0924-0136(98)00041-7

Mascarenhas, J., & Esperto, L. (2006). Direct Metal Laser Sintering (in Portuguese). Ciência & Tecnologia dos Materiais. *Journal of Portuguese Materials Society*, *18*(3/4).

Mehrabi, M. M Galip, G., Ulsoy, A., & Koren, Y. (2002). Manufacturing Systems and Their Design Principles, In Osita D.I. Nwokah and Yildirim Hurmuzlu (eds.), *The Mechanical systems design handbook: modeling, measurement, and control* (Vol. 1 – Manufacturing) Boca Raton, FL: CRC Press LLC.

Noguchi, L. D. (2003). *The Goldsmith and the Skill: The Jewel Quality on the Interface Between Design and Production Process in the Handmade Jewellery*. (Thesis). Federal University of Minas Gerais, Belo Horizonte, Brazil. [in Portuguese].

Nwokah, O. D. I., & Hurmuzlu, Y. (2002). *The Mechanical systems design handbook: modeling, measurement, and control*. Boca Raton: CRC Press LLC.

Osgrove, B. (2005). *Comparison of Low Cost Rapid Prototyping Machines*. Retrieved from rpworld.net

Ray, N. (2009). *Jewellery Manufacturing*. Retrieved February 22, 2009, from Jewel Info 4 U, www.jewelinfo4u.com

Richards, T. R., Lorrain, H., & Hilton, P. D. (2000). The Role of Rapid Tooling in Investment-Casting Applications. In Hilton PD. Jacobs PF (eds.), *Rapid Tooling: Technologies and Industrial Applications.* New York: Marcel Dekker.

Rocha, V. C. (2007). *The Use of Information Systems on Jewellery Design and Production Process*. (Thesis). Federal University of Minas Gerais, Belo Horizonte, Brazil. [in Portuguese].

Siu, N. W. C., & Dilnot, C. (2001). The challenge of the codification of tacit knowledge in designing and making: a case study of CAD systems in the Hong Kong jewellery industry. *Automation in Construction*, *10*, 701–714. doi:10.1016/S0926-5805(00)00088-1

Smith, P. G. (1998). Leverage the Cycle-Time Capability of Your Rapid Prototypes. *Time-Compression Technologies. North American Edition, 3*(7), 36-43.

Soo, S. C., Yu, K. M., & Chiu, W. K. (2006). Modeling and fabrication of artistic products based on IFS fractal representation. *Computer Aided Design*, *38*, 755–769. doi:10.1016/j.cad.2006.04.003

Stamati, V., & Fudos, I. (2005). A parametric feature-based CAD system for reproducing traditional pierced jewellery. *Computer Aided Design*, *37*, 431–449. doi:10.1016/j.cad.2004.07.004

Stamati, V., & Fudos, I. (2007). A feature based approach to re-engineering objects of freeform design by exploiting point cloud morphology In *Proceedings of SPM 2007: ACM Symposium on Solid and Physical Modeling*. Beijing, China.

Wong, T. N., Cheung, C. H., & Lau, H. (1999). Decision support system for a jewellery manufacturer. *International Journal of Production Economics*, *60-61*(20), 211–219. doi:10.1016/S0925-5273(98)00142-X

KEY TERMS AND DEFINITIONS

Computer Aided Design and Computer Aided Manufacturing (CAD/CAM): Systems are diverse software and hardware technologies, used in many applications in industries like automotive, aerospace, mechanical and others. One of the most important aspects of CAD/CAM implementation is the integration between different areas of enterprise, like product development, production and assembly.

Computer-Integrated Manufacturing (CIM): An engineering approach for manufacturing, where computer technology controls different functions in the enterprise. The use of CIM technology provides an effective communication among design, analysis, planning, purchasing, cost accounting, inventory control, and distribution. However, the CIM can be understood like a management strategy, using the computer resources to integrate functions and people in the corporation.

Fused Deposition Modeling (FDM): Process involves heating a filament of thermoplastic polymer and squeezing it out like toothpaste from a tube to form the Rapid Prototyping layers. The machines range from fast concept modellers to slower, high-precision machines. The materials include polyester, ABS, elastomers, and investment casting wax.

"Investment Casting" or "Lost-Wax Casting": also called by the French name of *cire perdue* or the Latin *cera perduta*, is the process by which a bronze, brass or other special alloys are cast from an artist's sculpture and jewels. The casting process is one of the oldest known metal-forming techniques, was first used during the period 4000-3000 B.C. Nowadays, with the appearance of the high-technology waxes, refractory materials and specialist alloys, the castings allow the production of components with accuracy, repeatability, versatility and integrity in a variety of metals and high-performance alloys, thus, in industrial uses, this modern process is called investment casting.

Mold: The exact and perfect replica of the piece to be made, copied either from a design or a piece/object. A mold is shaped around the shape/figure with the help of casting process. Molds are made from many materials, including plaster compounds. Some different methods of casting include the lost wax process, centrifugal or investment casting, and sand casting.

Rapid Manufacturing: A computer technology, the embodiment of the concept that allows defining elements or modules with well defined interfaces and "hooks" in a way that the process of manufacturing can be: developed in an independent manner, customized based on installation specific data, based on formal models of controller behaviour, amenable to automatic generation of control software modules, and reconfigurable to adapt to changes in the environment through changes to data and/or regeneration of control software.

Rapid Prototyping: A group of techniques, which has a vast application on several fields, such as architecture, art, engineering and industrial design. Using these technologies, and from a CAD3D geometrical file, it is possible in a few hours to build functional prototypes of components or even production tools, known as Rapid Tools or Rapid Manufacturing (RM).

This work was previously published in Handbook of Research on Trends in Product Design and Development: Technological and Organizational Perspectives, edited by Arlindo Silva and Ricardo Simoes, pp. 119-138, copyright 2011 by Business Science Reference (an imprint of IGI Global)

Chapter 47
Cases Illustrating Risks and Crisis Management

Simona Mihai Yiannaki
European University, Cyprus

ABSTRACT

Undoubtedly, the nature of the relationships between business and risk factors in one country or another does not fit exactly into a "model" nor does it have a pure placebo effect. Yet, models' simplicity may appeal to managers and regulators in understanding important business risks and crisis related phenomena. Backed by this idea, this research underpins a comparative study on SMEs handling risk and crisis management according to a new tailored model of a balance scorecard. This new model of a risk and crisis management aims at improving both SMEs management adaptation and performance across all of crisis' stages, something not attempted so far in the literature. The application of such a 'balanced-scorecard' comes from the author's experience as a banker financing various SMEs industries, as a bank consultant on risk management as well as primarily from the results of a survey performed on a sample of 48 Romanian and Cypriot SMEs, equal-proportionally selected from the area of trading, manufacturing, and services. The results of this case study show, coincidently or not, that there is a significant improvement of the financial performance of the SMEs that employed this model compared to those that did not. The monitoring period: 11/2009-06/2010 was employed as a representative one for the latest global financial crisis which affected the entire European Union region, as well.

INTRODUCTION

The last few years brought severe waves of globalisation and several financial and economic crises, where even large, strong, formerly sound corporations did not cope with. In particular, it seems like in the past 20 years there have been several so-called "once-in-a-lifetime events," such as the floods of Hurricane Katrina or any of the financial crisis, including 1987, 1998, 2000, or 2008, and "no one saw that coming." (Hubbard, 2009) Now it seems the moment to realize the need to protect the most 'numerous' of the firms

DOI: 10.4018/978-1-4666-1945-6.ch047

population: the small and medium size enterprises (SMEs). However, the contribution of SMEs varies considerably across countries. In most economies, the share of enterprises with less than 20 persons engaged is more than 70%, ranging between 67% in Ireland and above 95% in Greece. SMEs also account for a smaller share of the total number of employees, ranging between around 11% in the United States and the Czech Republic and more than 35% in Greece (OECD, 2010).

Since SMEs are much more exposed and fragile during and after financial than any other type of companies, whilst they are firmly identified by the EU as the core engine for European economic recovery, growth and innovation,, this chapter looks at a double loop feedback and SME stakeholders' benefit in implementing innovative risk management methods and techniques under a broad Business Socially Responsible strategy.

Subsequently it provides key elements of crisis management resolution not only in times of economic and financial distress but also in times of "black swan" events (Taleb, 2007), such as sudden drops of the National Stock Exchange Index, strong foreign exchange devaluations, sudden and acute raise in unemployment or any other temporary or chronic externality.

Although risk measurement can only be generated by historic data, it has been found that experienced estimators could, using PERT or other well-known models, create usable estimates, which gave usable and effective predictions, with computed standard deviations. One can measure and model risk whatever the risk, static and dynamic, however none can have a crystal ball, as one always have to deal with various levels of uncertainty. This uncertainty is sometimes called error. Any model has also at least 3 inner-errors: model error, parameter error and random error. The task is to minimize them accordingly, but never be able to eliminate them completely. Moreover, the problem of risk management arises further, when the firm doesn't consider any of these.

Others (for example Hubbard, 2010) consider that there are imaginary obstacles that must be overcome, as there are practical measurement solutions that can be applied to any uncertain decision. Hence, one can quantify any uncertainty and then compute the value of reducing that uncertainty by measurement. However one can see during the recent sub-prime crisis that even some of the best quantitative finance modelers at top rating agencies were not able to measure risks of structured products with complicated pay- off's.

Does the picture look though simpler or more effective for the SMEs or not? Except for paying attention at the error and uncertainty part, shouldn't one also look at what is inside the model, before starting measuring?

In pursuing risk measurement, one can learn about the revolutionary thinkers including John von Neumann (inventor of game theory), Harry Markowitz (grandfather of portfolio theory), and the late Fischer Black (Black Scholes option formula), or also find enlightening stories about Fibonacci numbers, chaos theory, the bell curve, regression to the mean, and more (Bernstein, 1996). Yet, despite all the intelligence, computer power, and sophisticated techniques, Bernstein presents us with the growing body of evidence discovered by researchers including the late Amos Tversky's cognitive biases and others who "reveals repeated patterns of irrationality, inconsistency, and incompetence in the ways human beings arrive at decisions and choices when faced with uncertainty." (Bernstein, 1996: pp.271).

In the attempt to model risk, the study employed conditions that can provide indicators of identifiable categories (the five pillars of our proposed model). Once one identifies the patterns indicating the category of risk associated, then one can try to predict when a categorized event might occur. Upon such predictability background, one can attempt to minimize the effects and maximize any recovery efforts. In quantum mechanics, the Heisenberg uncertainty principle states by precise inequalities that certain pairs of physical properties, such as position and momentum, cannot be simultaneously known to arbitrarily high precision. That is, the more precisely one property is measured, the less precisely the other can be measured. Is this paradox trying to teach us a lesson in managing risk and crisis?

The chapter attempts to use though, under a general simplified framework, the key risk evaluation framework which nowadays SME's managers and staff could utilize in reaching their strategic, operational, marketing and financial targets within or not a crisis context.

One big challenge of the study is that we drew under a synthetic five pillar system a balanced-scorecard (herewith BS) which not only can track dawn repeated risk/crisis events and measure their impacts, but it can accommodate further potential risks analysis and mitigation while leaving room for creative improvement at the same time.

In the view of this research, the SMEs are being treated just like any company which needs to support its long-term soundness and profitability in a socially responsible way.

In this light, this study has build a cross-sectoral pool of SMEs from various industries from a particular environment and tested their potential ability to alleviate risk impact and deal with crisis management by utilizing the BS.

BACKGROUND: METHOD ANALYSIS AND CHAPTER FRAMEWORK

Issues, Controversies, Problems: The Balanced Scorecard Concept

This section starts by responding to the question, as why one should utilize the balanced-scorecard (BS) instrument (Kaplan and Norton, 1996) at the base of this model?

Although, there are a few empirical studies linking the use of BS to better decision making or improved financial performance of companies, broadcast surveys of usage have difficulties in this respect, due to the wide variations in definition of 'what a BS is'. Single organization case studies suffer from the 'lack of a control' issue common to any study of organizational change - you don't know what the organization would have achieved if the change had not been made (Lingle & Schieman, 1996), so it is difficult to attribute changes observed over time to a single intervention (the BS), despite obtaining results which demonstrate the usefulness of the BS (Mooraj *et al*., 1999).

Another criticism among others (Schneiderman, 1999), usually from pundits and consultants, is that the BS does not provide a bottom line score or a unified view with clear recommendations: it is simply a list of metrics (Jensen, 2001). Typically however, the unanswered question relates to things outside the scope of BS itself (such as developing strategies) (Rohm, 2004), which in reality could have a decisive impact in the organization change-over and financial existence. For this reason, the suggested BS is designed to incorporate these aspects, under a flexible quantification.

On the other side of the propellers of the BS, according to Kurtzman (1997), it appears that 64% of companies questioned were measuring performance of companies in a similar way to the BS. Yet, again Kurtzman (1997) sees a strong controversy between two schools of thought. He refers to these schools as *the balance people* and *the numbers people*.

A synthetic table format of these two schools is presented in Table 1.

The balance people promote the use of corporate, or BS. Kurtzman utilizes the so called Analog's scorecard, which includes a number of

Table 1. The balance people and the numbers people (Kurtzman, 1997)

The Balance People	The Numbers People
The corporate scorecard should be balanced, including nonfinancial soft matters like employee satisfaction.	The balanced scorecard confuses the issue. Measurements should be purely financial.

nonfinancial measures such as number of new products, the rate of on-time deliveries, product development cycle times. He makes his scorecard link these measurements to financial indicators such as the percentage of sales increase due to new-product introductions or gross margins post introduction of new products.

Kurtzman uses Shell Oil as an example in the numbers people category, which seems to be a good example of a high potential pooled-risks affected corporation. Shell's business model is based on a matrix with four financial measurements as indicated in Table 2.

Shell management's view is that a balanced group of measurements with a number of 'soft categories' leads to confusions. This company's view is an explicit result driven one, where each activity is measured in terms of financial benefit, which in reality is neither a transparent way of sharing the company strategy nor it can differentiate between core activities and corporate social responsibility ones, for example.

At the same time, choosing to calculate a certain financial indicator over another seems to be more or less complicated or fashionable, for example, for finance people the EVA is considered a more accurate measure of control and profitability because it measures an opportunity cost to the money one is investing, and it might yield better returns elsewhere. The calculation allows for an adjustment to compare one project to a hypothetical alternative investment, as well as it is a short term related indicator.(Stewart, 1993). As long as it regards the bonus perspective, in order to be a successful tool, EVA based bonus systems should be long-term, based mainly on changes of EVA and offer considerable bonuses for considerable shareholder value improvements.

Table 2. Shell's business model- (adapted by the author)

Shell's Business Model	
Revenue Growth	Intrinsic Business Value (i.e., Company's Market value)
ROI (return on investment)	EVA (economic value add)

For SME's mangers, who most of the time being SME's owners of vartious background, the understanding of financial indicators takes time. However, more important for them nowadays is having to choose certain indicators over others which may make an important difference in the long run, as decision making is time bound. These schools of thought are putting together several controversial issues related to measurements in management and financial strategy, but they do not give a solution to what is best or why it is so.

Theoretically, Norton and Kaplan (1996) have used such methodology for large corporations not only for bringing in increased success and lessen managerial burdens, but also to reduce the agency conflict between mangers and shareholders (Jensen, 2001). In addition, this instrument will help more in understanding the strategic need for planning, measuring and having operational objectives for any SME, as well as linking budgets with strategy and HRM and constantly reviewing it, which defines the double loop orientation in risk mitigation and risk management.

Under this study, the aim is not the same, since in SME's managers happen to be the owners most of the time, while at the same time strong –cultural patterns may reveal differences in the way of doing business. Moreover, budget wise, SMEs are not as well equipped as large corporations to invest in all necessary risk packages or sustainable activities. First of all, this is due to environment culture, to lack of dedicated organizations and free funding for SMEs as well as lack of governmental support, in brief due to lack of regulation of this sector. Next by regulation, it is meant not necessary the control aspects, in the sense of fiscal or financial ones, but in the sense of business facilitators.

Therefore, this chapter focuses on channeling the attention of SME's management towards a combined 360degree model that does not have to be a burden in its application nor very time consuming in being utilized by the stakeholders involved. The novelty of this tool is that the BS is useful and interpretable not only by managers of SMEs in context, but also to the other stakeholders of the business for the purpose of transparency (www.bis.org).

Moreover, since risk management models are just like big weather forecasting models that predict financial storms, who may happen or not in the predicted timeframe, obviously, the models aren't always right. Even recognizing that the weatherman got it wrong a lot, it was helpful to know that it might rain tomorrow, because then one could make backup plans. This is not an excuse for any model's inaccuracy, but outlines the idea that some things are utterly immeasurable due to three common misconceptions, also called 'COM' (Hubbard, 2010), namely concept, object and method. A strong misconception about measurement is the objective of measurement itself. If "strategic alignment", or "employee empowerment" seem immeasurable, it is only because they are ambiguous. But if they are important to a business, then they must become measurable aims, and so they must be detectable in some manner, directly or indirectly (Hubbard, 2010).

To these, one may add perhaps the versatility of the risk mitigator and his/her interpretation altogether of the results.

Even by utilizing controlled experiments, variations on random sampling methods, and some very simple but non-obvious methods in many practical business situations, their interpretation is at stake when responding to problem solving. At the same time Hubbard (2009), looks at the way risk is measured by the four horseman, The four classes are actuaries, "war quants" economists, and management consultants. For that reason, it seems important to respond to the exact set audience, or have approached important aspects treasured by all these types of audiences. Hence, last, but not least, one could also add the calibration of subjective human inputs.

The Case Study: The Balanced Scorecard Designed for SME Risk Management

The main concerns of this study were if once utilizing the proposed BS model, the SMEs applying it could obtain any competitive advantage versus the ones not using it ? And, thereafter to see if there are any particular country differences?

The research had indeed used a relatively reduced sample and a relatively short period of time (12/2008-06/2009), considering that a crisis is based on three stages, pre-crisis, during-crisis and post crisis. Because of the pertaining status of the recent financial and economic crisis, it is still hazardous to estimate the completion time for the second and third phases. This is why the study can be expanded for a future larger sample for various other countries or regions, as well as considering a lengthier timeframe.

At the same time, as many of the SMEs are operating cross-industry, the sample chosen includes SMEs categorized based on having over 60% of their activity in either trading, manufacturing or services. In the sample, only about 35% of SMEs were operating solely in one of these industries. This is the reason why the research has concentrated the study's results on the overall response of the SMEs.

Starting from the debate of people vs. numbers the proposed BS has been designed under a five pillars system. The main rule SME's directors were given to follow was that risks can be better comprehended by the people/units most capable of understanding them, and then the SME's strategy and decision making may alter or not the existing risk metrics system.

It is important to mention that the results obtained were after the application of all five pillars at the same time and the results collection and interpretation was performed after the end of June 2010. Also, this means that all SMEs applying the BS have implemented during the analysis' period a risk management structure and have designed their own strategy and governance to follow the BS.

The figure next, represents the Conceptual Framework of the study under the BS concept, where the BS elements emerged are based on preliminary discussions with SME managing directors, from the literature review, as well as from the author's experience while dealing with SMEs financing over the past 10 years. Five aspects appeared to be essential in the risk assessment and measurement and they are presented under the following cluster:

Next, the study presents and analyses the five pillars characteristics aiming at integrating them under a Risk Management and Control Centre BS.

Pillar I: The Enterprise Responsible Package (ERP)

The ERP measures:

1. The degree of awareness of the need for having socially responsible stakeholders
2. The ability of the SME to connect and network at all levels with its stakeholders (starting first with the government and corporations and then with customers, employees and eventually with non-trading customers, potential future stakeholders (Jensen, 2001))
3. The ethical response of the SME towards the society
4. The protection of the environment, meaning how green the SME can become in the limits of its business and beyond.
5. The relationship management develops with all stakeholders, including customers

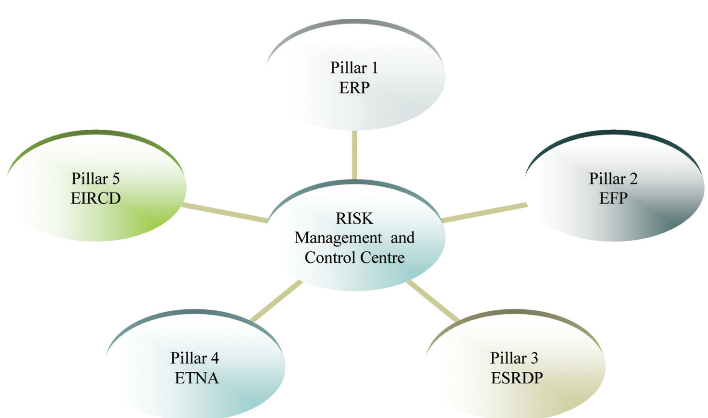
This pillar focus is mainly on measuring: customers' satisfaction, asset turnover, number of green activities per entity, market share increase, operational excellence.

The initiatives and spin-offs brought forth by the SMEs applying the offered risk management tool were: cross-industry partnerships, mergers and acquisitions, managerial and staff training, customers' loyalty schemes and development of corporate social responsibility actions, reinforcing SME-lender relationship (Collier, 2009).

Regarding strategy related aspects of this pillar, a broad outline of risk management at this first level has encompassed:

- Organisational structure definition and implementation or redesign;
- Comprehensive risk management approach;
- Risk management policies approved by the Board or the people in charge which should be consistent with the broader busi-

Figure 1. The conceptual framework of the proposed BS for SMEs



ness strategies, capital strength, management expertise and the overall willingness to assume risk;

- Guidelines and other parameters used to govern risk taking including detailed structure of prudential limits or thresholds of risk for each identified category (Glantz & Mun, 2008);
- Developing a strong Management Information System (MIS) for reporting, monitoring and controlling risks;
- Well laid down procedures, effective control and comprehensive risk reporting framework (Collier, 2009);
- Separate the risk management framework and team/designated body, independent of any operational department and with clear delineation of levels of responsibility for management of risk;
- Periodical review and evaluation not only for risk levels on short term, but also for policies and procedure on a medium to longer term; and
- Redesign checklists – highlighting updated key industry-specific hazards and control measures (Welham, 2003).

Eventually, the results obtained for the Pillar I are those obtained after the application of all the five Pillars of the balanced-scorecard are presented below:

Table 3 interpretations require attention to the major differences obtained, firstly in the area of increased customer satisfaction, which could translate in a serious image change in the eye of the customer for those SMEs approaching the risk management platform. Secondly, the area of green activities seems to have similar high impact in both countries. The audit evaluations eventually seem to have the strongest impact for both countries and in all sectors, mainly because of the international exposure of all of the SMEs within the sample, but also to the strict correlation to the financial success (that is evaluated next in Pillar II).

Table 3. Pillar I: Results per country

Pillar I Parameters Available	**Equivalent Bilaterally set Benchmarks**	**No application of risk management balanced-scorecard**		**Direct application of risk management balanced-scorecard**	
		Romania SMEs	Cyprus SMEs	Romania SMEs	Cyprus SMEs
a. customers' satisfaction increase	50% for all industries	-10%, -20%, and -22% respectively	-12%, -16%, and -22% respectively	60%, 35%, and 38% respectively	55%, 45%, and 18% respectively
b. asset turnover increase	30% for trading, 30% for manufacturing 40% for services.	0%, 0% -10%	10%, 15% 20%	20%, 25%, 28%	35%, 30%, 45%
c. number of new green activities per SME	10	0	2	5	6
d. raise in the market share	30% for trading, 10% for manufacturing and 20% for services.	Difficult to evaluate per SMEs	Difficult to evaluate per SMEs	Difficult to evaluate per SMEs	Difficult to evaluate per SMEs
e. operational excellence grade (Audit evaluations)	Maximum score 10 out of 10.	-1, 2 3	4, 5, 7	7, 6, 7	8, 8, 9

Pillar II: The Enterprise Financial Package (EFP)

Thanks to the simple financial structure, SMEs can respond faster to changing economic conditions and meet local customers' needs, growing sometimes in their size or becoming bankrupt. Here the study analyzes the set of financial indicators that have to be risk associated and treated for proper adequacy. The reason the research introduced this aspect is also due to finding the response to the Kaplan and Norton (1996) question: How should we appear to our shareholders?

In the original Z-score model of Altman (1968) used five out of 22 ratios as decisive in predicting bankruptcy: Working Capital/Total Assets, Retained Earnings/Total Assets, EBIT/Total Assets, Market Value Equity/Book Value of Total Debt and Sales/Total assets.

Despite having many supporters for the application of Basle II Accord to SMEs (Altman & Sabato, 2005), many criticism have been raised by governments and SME associations that high capital charges for SMEs brings closer credit rationing and could reduce economic growth.

Since banks are the main source of finance for SMEs, and since the Basel II accord allow banks to classify SMEs with turnover under EURO 50 million as either retail or corporate for capital requirements calculation, the banks will tend to treat SMEs under a more automated decision methodology, by scoring their financial ratios.

Several studies show that when banks classify SMEs as retail they obtain considerably lower level of capital adequacy than 8%, (www.bis.org) (i.e 4.76% when utilizing Z-Score models), while if they classify them as corporate, they obtain above 8%, (i.e. 8.6% for Z-score model) (Altman & Sabato, 2005). Why is this important? For this reason, banks who tend to split their SME portfolio in 50/50 retail and corporate, will not necessarily win in terms of risk mitigation and they will tend to develop customized risk prediction models for SME, for which reason, financial ratios of SMEs will gain further importance in acquiring bank funding.

However, the study considers only a small package of financial indicators, which are more appropriate to SMEs: Cash ratio (or the company liquidity position), Return on Investment (ROI) (The higher the ROI, the better the performance, and a higher ratio will also improve the firm's ability to find new investors), Return on equity (ROE) and Economic Value Added (EVA) and a Stress Tested Cash Flow ratio (Stress Tested Cash Flow/ Debt Payment+ Preference Dividend+ Interest).

However the example below may show that ROI and EVA results may be contradictory and more accurate interpretation is required.

Let's suppose of a SBU earning currently a return (ROI, ROIC,ROCE) of 30% and that this firm faces an investment opportunity producing a return of 20%

What happens to the ROI of the SBU if the investment is executed?

Before investment: Capital 100, Operating profit 30 or (NOPAT), Capital cost 10%

ROI = 30/100 = 30%, EVA = 30 - (10% x 100) = 20

Investment's capital requirement 20, return 20%/ year: Thus increase in yearly operating profit is 20% x 20 = 4.

After investment: Capital 120, Operating profit 34, Capital cost 10%

ROI = 34/120 = 28%, EVA = 34 - (10% x 120) = 22

In this case decreasing ROI is good for the shareholders, thus ROI should not be maximised and therefore it is problematic controlling tool. Usually large corporations have at least some very profitable units and particularly these units are steered wrongly with ROI.

Due to these contradictions, and for the purpose of having consistent and comparable data and results for interpretation, our recommendation to each SME applying the BS was to stick to its banker's set of financial indicators, as well as to calculate them just like their banker (example: yearend vs. average method of computation, same formulas and periods) (Schoenebeck and Holtzman,2010).

Although ratio analysis is a powerful tool in credit risk evaluation (Bacidore *et al.,* 1997), banks have understood that it is not the only means in this process, as they are other as important parameters for such assessment. Besides, high quality of one ratio may mean a poor quality of another dimension or ratio, as in the box above, also it may not guarantee a sustainable continuous basis of the respective ratio, as they are based on the past.

The results obtained after the application of all the five Pillars of the balanced-scorecard were as presented below and they may represent on average about 20% of the overall credit risk evaluation, where each ratio may have equal strength. The limitation here resides in the calculation of these ratios for the end period and comparison with only the same period a year before not with a trend of at least 5 years.

Last, but not least important, the study employed the stress test of the cash flow from operations, which proved to be also a sensitive issue for the SMEs in Romania, but less sensitive for those in Cyprus. The positive cash flow indicator shows the ability of the SME to satisfy the potential cash flow constraints in case of credit default. At the same time, it is a clear cut indicator compared to meeting just industry adjusted liquidity levels. Although there is no benchmark for the volume, an arbitrary level of 20% was set, which was met by all SMEs applying the risk BS, but not fully by those not applying our risk BS. This may be correlated to the positive results of the other financial indicators, but it proved to be a rather sensitive indicator in this case.

The financial improvement for the SMEs utilizing the risk BS proposed seems overall higher for both countries, which may be or not related to the parallel implementation of the other Pillars' elements, or just as a placebo effect or a noise effect in the market.

The results for these ratios were obtained directly from the SMEs, after being provided with the same formula of calculation. Obviously the element of trust was essential, but also the double calculation based on the companies' key financial statements. However, due to differences in accounting standards and taxation, there are always differences in the input results, hence in the final results. However, since 60% of the firms were medium size companies, they have adopted IAS and therefore the margin of error has significantly diminished, being left only with the input error probabilities. However, to understand risk, one needs to understand the distribution of potential

Table 4. Pillar II: Results per country

Pillar II Parameters Available	Equivalent Bilaterally set Benchmarks	No application of risk management balanced-scorecard		Direct application of risk management balanced-scorecard	
		Romania SMEs	Cyprus SMEs	Romania SMEs	Cyprus SMEs
Cash ratio	0.5 to 1	0.15	0.25	0.35	0.44
ROI	15%	3%	5%	14%	12%
ROE	15%	8%	10%	16%	12%
EVA	15%	2%	4%	12%	10%
Stress Tested Cash Flow	Depends on industry, set up at 20%	No	50% No	Yes	Yes

outcomes and also apply consistently the same calculation/computation formula.

Consequently, it seems there is a definite opportunity to take actuarial type standards to modeling insurance risk for example, and apply them to other operational, financial and strategic risks that organizations face.

As for the stress tests, (Chorafas, 2006) covering stress analysis and the use of scenarios, models, drills, Monte-Carlo simulation, benchmarking, back-testing, and post-mortems, creditworthiness, wrong way risk and statistical inference, probability of default, loss given default and exposure at default, stress testing expected losses, correlation coefficients, and unexpected losses, stress testing related to market discipline and control action, and pillars 2 and 3 of Basel II (www.bis.org)., all these seem to be very complicated to an SMEs, therefore they find their usefulness in banks and the financial sector and rarely in SME sector. For such a reason, the levels were bilaterally settled. This is not only due to their intricate nature and data warehouse needs, but also due to the human factor interpretation or, moreover, response in the business world.

Further stress tests parameters could be added for SMEs performing import-export activities, in order to observe and control associated foreign exchange exposures (Aabo, 2010) and hedging techniques (Woods *et. al.,* 2007).

Pillar III: The Enterprise Supervision and Resource Development Package (ESRDP)

This is a redesigned part of the BS that looks into the development of the learning and growth aspects of the HRM, looking at: employee turnover, investment rate, illness rate, performance pay, gender ratios and frequency of supervision.

Despite arrangements to follow a risk management procedure this part of the BS has encountered difficulties in implementation due to the low level of employment of several firms from our sample, hence results may get a rather poor interpretation.

The only ratios available for Pillar III in this situation are the following:

Before assessing the above indicators the firms applying risk management have acquired several series of knowledge transfer regarding the guidelines for risk management implementation.

There is no clear trend in this series over the analyzed period, except for the decrease of employee turnover for the case of Romania, perhaps due to less people finding opportunities for employment during financial crisis in the SME sector. Interestingly, there is a slight increase of the female employed population versus the male one in the SMEs applying risk BS in both countries, compared to those who do not. But, there is quite a dramatic tendency for unemployment to rise sharply during periods characterized as economic recessions. Both countries have experienced a

Table 5. Pillar III: Results per country

Pillar III. Parameters Available	Equivalent Bilaterally set Benchmarks	No application of risk management balanced-scorecard		Direct application of risk management balanced-scorecard	
		Romania SMEs	Cyprus SMEs	Romania SMEs	Cyprus SMEs
Employee turnover	n/a	20%	5%	5%	4%
Gender ratio	n/a	38% women	30% women	41% women	33% women
Frequency of Supervision	n/a	The same	The same	Increased control and audit	Increased control and audit

superior/effective application of the control and audit function. Indubitably, this last parameter of Pillar III is correlated not only with the other financial and ERP applications of the proposed BS, but also with the way the management empower personnel and allocate the bonus system. Some results also suggest that a firm's incentive or ability to offer fair employee treatment is an important determinant of its financing policy. (Bae, *et al.,*2010).

Also, the relation between incentives and risk is more positive when the manager's effort is more effective in collecting information or in acting upon it. Using data on chief executive officers (CEOs), Shi (2011) found that incentives for CEOs increase with industry-wide risk, a measure of respondable risk. The positive relation diminishes when the CEO is less able to collect information or is less effective in acting upon it. Reason, for which we recommend even more a set up of a tailored scorecard for the Pillar III.

If at bank level, Cash *et. al.* (2010) found only modest evidence that CEO compensation plans promote significant firm-specific heterogeneity in bank risk measures or risky activities, for SME's executives, due to the nature of the activity, this may not be valid.

In this case a SME's general manger or managing director would be in charge to appoint someone to lay down risk parameters and establish risk management control systems. This stage includes the following the below steps:

- Risk Management Architecture / Structure for all risk areas encountered by the industry and the particular firm.
- SME's to get the necessary equipment in getting the ability to identify, measure, monitor and control the various risks.
- Set up an Integrated Risk Management Framework, such as the study's BS or a similar system.
- Move towards Risk based Supervision and Risk focused Internal Audit to enhance risk management functions.
- Set up a Change Management Team or a Crisis Management Team.
- Perform a Risk aggregation and budget allocation for this function and its activities.

Pillar IV: The Enterprise Technological Networking Advance (ETNA)

This particular pillar looks at the technological environment impact on the business and measures indicators such as: level of computerization, number of software packages, number of technological changes in a year time, return on investment in technology.

Regarding ETNA, most SMEs need to build up an Industry Databases which makes available information from an organised centre that offers control for offices and operating levels and help in the risk analysis and in the performance of daily activities.

Table 6. Pillar IV: Results per country

Pillar IV. Parameters available	**Equivalent Bilaterally set Benchmarks**	**No application of risk management balanced-scorecard**		**Direct application of risk management balanced-scorecard**	
		Romania SMEs	Cyprus SMEs	Romania SMEs	Cyprus SMEs
Level of computerization	n/a	The same	Average increase of 2 computers per firm	Average increase of 4 computers per firm	Average increase of 5.5 computers per firm
Software packages	n/a	The same	The same	10% increase	25% increase
ROI in technology	n/a	n/a	n/a	n/a	18% increase

The degree of computerization of the two countries at the sample level have increased, with a larger spur in the Cypriot firms on both hardware and software, and we assume that this was reflected also in the ROI in technology figures. For certain groups of companies operating in manufacturing investment in technology seem to be benefic also in financial terms even on a short term period in Cyprus, while such information is only assumed in Romania, too. In this sense, more explicit details should be obtain as to the purpose of the acquired technology, the business cycle stage of the firm (Hamilton, 2010), if for incremental change, for improvement of daily activity, or for radical change (as in Pillar V), or for facing market demand shifts (despite contraction of demand), or even for future strategic plans.

Pillar V: The Enterprise Innovation and Radical Change Development (EIRCD)

At this stage we assume already that the SME will be exposed to previous learning methods and patterns and we try to identify the possible risk related to implementation of new change processes due to either risks or economic/financial crises. The indicators which may measure this development stage are: employee turnover, investment in innovation, number of radical changes in a year regarding strategy, operational profit margin change post innovation and product leadership per innovation.

This last pillar suggests a high distribution of innovation at the level of SME for both countries, with present both incremental and radical innovations, and with almost double expenditure in Cyprus, but with more results in Romania, although the Cypriot innovations seem to be more profitable. These innovations may account more or less for the employee turnover reduction, where investment in such areas brought a higher stability on freeze of staff on the market, but are certainly provided a cutting edge in the future competitiveness of these firms.

Solutions and Recommendations Based on Discussion of Results

If we consider our model based on a general linear model: Yp=∑ (XB)+εi, where Yp is the study's risk metrics final result or the response variable of each SME applying the suggested model, ∑ represents the sum from 1 to 5 of each pillar's result X, which is the a random discrete variable of almost sure equality that belongs to {0,100} or has an alphanumeric equivalent translated into

Table 7. Pillar V: Results per country

Pillar V. Parameters Available	Equivalent Bilaterally set Benchmarks	No application of risk management balanced-scorecard		Direct application of risk management balanced-scorecard	
		Romania SMEs	Cyprus SMEs	Romania SMEs	Cyprus SMEs
Employee turnover	n/a	20%	5%	5%	4%
Investment in innovation	EUR 50k/ firm	n/a	n/a	EUR20k/firm	EUR40K/firm
Number of radical changes regarding strategy	5/5	0	0	4/5	3/5
Operational profit margin post innovation	20%	n/a	n/a	16% manufacturing 20% services	18% manufacturing 18% services
Product leadership per innovation	Minimum 1 per county	n/a	n/a	2 (service industry)	0

a numerical range belonging to {0,100}, B is a fixed parameter of 0.2 (the equivalent of a equally distributed weight of each pillar), and εi is an unobserved variable that adds noise to the linear relationship between the dependent variable and regressors (X), which for each case we assume it to be close to zero. Eventually the maximum score of the BS will be Yp=100x0.2+100x0.2+100x0.2+100x0.2+100x0.2+0=100.

Where there is no value of the components that make up X, 100 points will be divided for the filled in values only, thus their weight in X become slightly higher.

However, for better interpretation of results, a larger sample together with a more detailed separation of the SMEs according to their representative industries is recommended.

A unique characteristic of this BS is that it can be customized to the specific needs of the company, industry, business cycle, manger and stakeholders, depending also on the emerging risk assessment results and on the capacity to interpret risk of the organisation. This way, the BS will offer the flexibility for the SME's management own judgment and appraisal of the firm's generic and specific risks, as well as offering an adaptable tool to the market change and the technology change in risk management techniques.

A simple Final Rating Model of the BS suggested for SMEs, could look like Table 8.

There is always room for providing a minimum passing grade for each pillar (perhaps 15, depending on industry dynamics) and for the entire scorecard (perhaps 80 instead of 100). However, after getting the score most important becomes the choice of the most appropriate solutions in risk management.

Therefore, in view of the presented results in the above section, one should interpret all these result by correlating both the pillars among them and the indicators within each pillar.

For example on Pillar I, the two countries SMEs obtained mixed results on customer satisfaction on both categories of samples (with no risk and direct risk management application). Although, the results show that Romanian SMEs who have applied risk management as per this study guidelines have obtained higher customer satisfaction on average than the Cypriot SMEs which did so, these results create a paradox interpretation. At the same time we need to understand the results of the following pillars to see if this is a long term or an erratic change.

As for the ratios calculated as per Pillar II, one can notice an improvement of liquidity ratios for the SMEs with risk management in both countries, registration of higher standard deviation for Romania for the SMEs with risk management tools.

This may indicate a higher concentration of risk on a particular industry (manufacturing for Romania) or a different application style and

Table 8. The Balanced Scorecard final rating and interpretation

Pillars	Pillar I	Pillar II	Pillar III	Pillar IV	Pillar V
Overall Score	100	100	100	100	100
Obtained Score	=90x0.2+0= 18	=80x0.2+0=16	=75x0.2+0=15	=50x0.2+0=10	=75x0.2+0= 15
Areas of risk mitigation (example)	Insufficient green activities	Insufficient ROI and EVA	High employee turnover	Identify benchmarks	Increase investment in innovation
Recommendations	Identify ways of implementing more green activities	Seek alternative sources of capital,	Develop or change the performance pay system	Seek industry Database and build up the MIS, perform data mining.	Supervise innovation output and return on investment in technology.

management rigor of the scorecard. The results show also a significant change from inefficient investments to efficient ones, especially in the area of manufacturing, and especially for Romania, and at a lesser extend for Cyprus, where investments in trading seem to be more efficient. These results have a standard deviation at a constant average of 8%, while highest average is obtained in manufacturing in Romania and in services and manufacturing in Cyprus.

Again, Romania sees to obtain the highest value of increase in EVA, perhaps due to the ability to pass through the financial crisis.

However, when correlating ROI, EVA and operating profit results for Romania, it appears that the cost of capital is dearer in this country. At the same time the industry with a higher increase is again the manufacturing one, followed by trading one.

When interpreting the results of Table 9, regarding the four financial ratios calculated for the two types of SMEs in the two countries, one can notice the same standard deviation of EVA as for the ROI for SMEs applying our BS risk model in Romania, which is consistent to the two indicators' controversy presented in section 1, Pillar II. Other discrepancies are seen in random higher standard deviations for the SME which applied our BS, particularly in Romania, except for the case of the ROE indicator, where the standard deviation is higher in the case of SMEs applying no-risk model. The same statistical tool shows relatively lower values variation for the SMEs in Cyprus applying the BS, as well as smaller differences in the average values between SMEs applying the BS and those who did not, compared to the Romanian counterparts.

For Cyprus, the highest spur of average values is obtained by SMEs applying risk management tools in the service industry. Also, for Cyprus there is a higher increase in the average ROI vs. the increase of the average EVA for SMEs applying BS. This could be related to the lower ROE and the higher increase in capital for the Cypriot companies for capital expenditure on innovation, radical change and MIS or on computer and technology acquisitions.

The Cash Flow Stress Test shows the ability of SMEs to pass through liquidity problems. Both countries' SMEs which applied risk management scorecard have passed this test. Those SMEs who did not apply the BS failed the test completely in Romania, whilst only 50% of the Cypriot SMEs non BS risk applicant passed the test. This may be correlated to different market mechanisms and different funding facilities available for SMEs in the two analysed countries, or to each company's economic strength, or to the support of the findings of the literature review (Shi, 2011) regarding executive mangers' compensation packages based on their ability to mitigate risks.

Table 9. Pillar II's ratio averages and standard deviations for the cash ratio, ROI, ROE and EVA

	Cash Ratio		**ROI**		**ROE**		**EVA**	
	Average	STDEV	Average	STDEV	Average	STDEV	Average	STDEV
SME (Romania non risk)	0.15	0.07	0.03	0.05	0.08	0.08	0.02	0.04
SME (Romania risk BS)	0.35	0.10	0.14	0.06	0.16	0.06	0.12	0.06
SME (Cyprus non risk)	0.25	0.09	0.05	0.07	0.10	0.08	0.04	0.06
SME (Cyprus risk BS)	0.44	0.05	0.12	0.05	0.12	0.08	0.10	0.06

In the same manner one can analyse the qualitative indicators, paying attention at meeting the benchmarks, however, there is no exact recipe for getting the best value since one high indicator may mean a possible negative response in another indicator.

Obviously in the end there is a set of the questions that sends us back to Hubbard's literature review of whether we have chosen the right indicators or ratios and if we have calculated them consistently, if the interpretation is the same or if the strategic objectives are met or not eventually. For this issue the fact that the SME develops procedures and involves cross-departmental teams in the area of risk management proves to be an additional risk mitigator, if not even a performance enhancer.

FUTURE RESEARCH DIRECTIONS

It is worth trying to make an analysis as a separate research that crosses the borders of a case study and adds in a large number of SMEs from various countries or regions and upon a larger number of defined indicators to be able to draw a more accurate combination of ratios that are more appropriate in risk evaluation and mitigation for the future generation of SMEs.

Another issue is related to the potential cultural aspects in the way risks are dealt with or risk management is applied in general, the level of uncertainty acceptance, the entrepreneurial spirit in each country, as well as the trends regarding particular reasons for defaulting in one country or another, if any. This may be related to other structures of externalities beside the common affecting ones, such as corruption levels, red-tape level, political implication of actions and others.

From the point of view of comparison against benchmarks or benchmarks translation may be required. It could be mentioned also that a better approach may to set benchmarks not only for periods of boom, normal growth, but also for recession times.

This approach has not been used within this study, the benchmarks being commonly agreed with all the participating subjects (i.e. the managers of the SMEs' sample).

Regarding the fine tuning of the ratios calculation or even choice or the sets of qualitative indicators, the advantage of this balanced-scorecard is that it provides room for adaptation to any industry and it shows how to use and measure or put together under a flexible rating system the obtained results.

CONCLUSION

Upon the analysis of a representative number of SMEs in Cyprus and Romania in the area of trading, manufacturing and services, the findings show the areas where the two countries provide stronger versus weaker appraisal indicators. The strong areas for Romania seem to be manufacturing and trading, while for Cyprus the service industry, all these neglecting the level of the cost of capital in both countries.

The BS utilised indicators have been analyzed from both quantitative and qualitative perspectives and could provide a feedback loop for improvement of SMEs activities as well as for paying attention to other uncovered areas of risk management, such as measuring the impact of a "once in a lifetime" event.

If risk management programs really do work, then it seems logical to assume that companies in a given industry with a (self proclaimed) "highly effective" risk management program would show greater shareholder returns, less earnings volatility, and better safety and regulatory compliance records (Mihai-Yiannaki, 2010) than other companies in their peer group who lack such a program. Yet, there appears to be no valid evidence that current risk management practices, taken as a whole, serve to improve overall corporate performance. The evidence just isn't there, and the research thus has to agree with Hubbard (2010).

The calibration on the rating scale or a balanced-scorecard for SMEs is expected to define also the future deals pricing and other related terms and conditions for the undertaken transaction or operational exposures. Also by tracking down previous transactions' risks, any SME can find resources to discover new business relationships and/or deals, or to improve existing ones, renegotiate or give up certain business relationship or even take legal actions in due time.

Such a prediction or simulation is important not only for forecasting and planning purposes in operational management of the firm, but for accounting and provisioning purposes, as well as for future capital budgeting or even for strategy redesign and resources redeployment.

Moreover, by having such a risk management system in place in parallel with a sound firm strategy, the SME becomes more equipped for the future, its management and human resources more astute on risk mitigation techniques, overall being capable to control better the company and to adapt to crises conditions and adverse risks, altogether creating a cultural-cognitive pressure for environmental sustainability.

Whatever measurement of risk derived out of models especially in operational risks, it is difficult to do back- testing or stress testing. The measurement may be correct (not the absolute figure) in some cases for predicting the trend if other things remain constant. Hence, measurement of tail events does not reflect the true picture of uncertainty in short term (within a year). Therefore, measurement of risks shall provide some strategic inputs towards the direction/movement of risks not in absolute terms.

Thus, it's not just a case of accurately assessing and management individual risks, associated to each SME's activity but also in considering the extent to which there might be a "domino" or "contagion" effect among different risk factors, or among the risk pillars, and how can the SME interpret and integrate the results.

Eventually, this study does show significant improvement in the results obtained by risk management lead SMEs in both countries, compared to those not operating by this norm.

REFERENCES

Aabo, T., Høg, E., & Kuhn, J. (2010). Integrated foreign exchange risk management: The role of import in medium-sized manufacturing firms. *Journal of Multinational Financial Management*, *20*(4-5). doi:10.1016/j.mulfin.2010.08.002

Altman, E. I., & Sabato, G. (2005). *Modeling credit risk for SMEs: Evidence from the US market* (December 26, 2005). Retrieved June 2010, from http://ssrn.com/abstract=872336

Altman, E. I., & Sabato, G. (2010). Effects of the new Basel capital accord on bank capital requirement for SMEs. *Journal of Financial Services Research,* Forthcoming. Retrieved June 2010 from http://ssrn.com/abstract=707204

Bacidore, J. M., Boquist, J. A., Milbourn, T. T., & Thakor, A. V. (1997). The search for the best financial performance measure. *Financial Analysts Journal*, (May/June): 11–20. doi:10.2469/faj.v53.n3.2081

Bae, K. H., Kang, J. K., & Wang, J. (2010). Employee treatment and firm leverage: A test of the stakeholders theory of capital structure. *Journal of Financial Economics, 100*(1). Elsevier Inc. doi:10.1016/j.jfineco.2010.10.019

Basel Committee on Banking Supervision. (2003). *The new Basel capital accord.* Retrieved July 2010 from http://www.bis.org

Basel Committee on Banking Supervision. (2004). *International convergence of the capital measurement and capital standards.* Retrieved June 2009 from www.bis.org

Bernstein, P. L. (1996). *Against the gods: The remarkable story of risk. Kindle Edition*. Wiley.

Cash Acrey, J., McCumbera, W. R., & Nguyena, T. H. T. (2010). CEO Incentives and bank risk. Journal of Economics and Business, 63(5). Elsevier. doi:10.1016/j.jeconbus.2010.09.002

Chorafas, D. (2006). *Stress testing for risk control under Basel II*. Butterworth Heinemann.

Collier, P. M. (2009). *Fundamentals of risk management for accountants and managers*. Butterworth Heinemann.

EC. (2003). Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. Retrieved June 2009 from http://www.euresearch.ch/fileadmin/documents/PdfDocuments/Micro_Small_Medium_Enterprise_-_EU_Definition.pdf

Glantz, M., & Mun, J. (2008). *The banker's handbook on credit risk.* Academic Press. Hamilton, J. D. (2010). Calling recession in real time. *International Journal of Forecasting, 27*(4). International Institute of Forecasters. Elsevier B.V. doi:10.1016/j.ijforecast.2010.09.001

Hubbard, D. W. (2009). *The failure of risk management: Why it's broken and how to fix it*. Wiley.

Hubbard, D. W. (2010). *How to measure anything: Finding the value of intangibles in business*. Wiley.

Jensen, M. C. (2001). Value maximization, stakeholder theory, and the corporate objective function. In Andriof, J. (Eds.), *Unfolding stakeholder thinking, Greenleaf Publishing, 2002*.

Kaplan, R. S., & Norton, D. P. (1996). Using the Balanced Scorecard as a strategic management system. *Harvard Business Review*, 76.

Kaplan, R. S., & Norton, D. P. (1996b). *Balanced Scorecard: Translating strategy into action*. Harvard Business School Press.

Kurtzman, J. (1997, February 17). Is your company off course? Now you can find out why. *Fortune*, 128–130.

Lingle, J. H., & Schieman, W. A. (1996). From Balanced Scorecard to strategic gauges: Is measurement worth it. *Management Review*, 85.

Loader, D. (2006). *Operations risk. Managing a key component of operational risk*. Butterworth Heinemann.

Mihai-Yiannaki, S. (2010). Bank shares restore under recent industry bailout. *World Review of Entrepreneurship* [Inderscience Publishers.]. *Management and Sustainable Development*, *6*(4), 336–349.

Mooraj, S., Oyon, D., & Hostettler, D. (1999). The Balanced Scorecard: A necessary good or an unnecessary evil? *European Management Journal*, *17*(5). doi:10.1016/S0263-2373(99)00034-1

OECD. (2010). *SMEs, entrepreneurship and innovation*. Paris, France: OECD.

Ong, M. (2005). [Academic Press]. *Risk Management*, *1*,.

Rohm, H (2004). A balancing act. *Perform Magazine, 2*(2).

Schneiderman, A. M. (1999). Why Balanced Scorecards fail. *Journal of Strategic Performance Measurement, 6*.

Schoenebeck, K., & Holtzman, M. (2010). *Interpreting and analyzing financial statements* (5th ed.). Pearson.

Shi, L. (2011). Respondable risk and incentives for CEOs: The role of information-collection and decision-making. *Journal of Corporate Finance*, *17*(1), 189–205..doi:10.1016/j.jcorpfin.2010.10.003

Spedding, L. S., & Rose, A. (2007). *Business risk management handbook*. CIMA Publishing.

Stewart, G. B. (1993). EVA: Fact and fantasy. *Journal of Applied Corporate Finance*, *7*, 6–19.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. Random House.

Todinov, M. (2006). *Risk-based reliability analysis and generic principles for risk reduction*. Elsevier.

Welham, J. (2003). *Tolley's risk assessment workbook series: Retail*. Butterworth Heinemann.

Wendling, C. (2010). Explaining the emergence of different European Union crisis and emergency management structures. *Journal of Contingencies and Crisis Management*, 18(2), 74-82. Blackwell Publishing Ltd. DOI: 10.1111/j.1468-5973.2010.00606.x

Woods, M., Kajuter, P., & Linsley, P. (2007). *International risk management*. CIMA Publishing.

ADDITIONAL READING

Berge, T. J., & Jordà, O.̀ (2009). The classification of economic activity. *Working paper*, University of California, Davis.

Camacho and Perez-Quiros. (2010). 2010 M. Camacho and G. Perez-Quiros, Introducing the EURO-STING: short term indicator of Euro area growth. *Journal of Applied Econometrics*, *25*, 663–694. doi:10.1002/jae.1174

Chauvet, M. (2010). The beginning and end of the 2007–2009 recession. sites.google.com/site/crefcus/probabilities-of-recession/The-beginning-and-end-of-the-2007-2009-recession1, April 2010.

Chauvet, M., & Potter, S. (2010). Monitoring business cycles with structural changes. *International Journal of Forecasting*, *26*(4), 777–793. doi:10.1016/j.ijforecast.2009.08.003

Croushore and Stark. (2003). 2003 D. Croushore and T. Stark, A real-time data set for macroeconomists: does the data vintage matter? *The Review of Economics and Statistics*, *85*, 605–617.

Dueker, M. (2008). Predicting the trough and a jobless recovery. www.econbrowser.com/archives/2008/12/predicting_the_1.html. December 9, 2008.

Estrella, A., & Mishkin, F. S. (1998). Predicting US recessions: financial variables as leading indicators. *The Review of Economics and Statistics*, *80*, 45–61. doi:10.1162/003465398557320

Gilchrist, (2009). 2009 S. Gilchrist, V. Yankov and E. Zakrajšek, Credit market shocks and economic fluctuations: evidence from corporate bond and stock markets. *Journal of Monetary Economics*, *56*, 471–493. doi:10.1016/j.jmoneco.2009.03.017

Hamilton, J. D., & Owyang, M. T. (2009). The propagation of regional recessions. W*orking paper*, UCSD.

Importance of Basel II Regulation/Supervision. (October 17, 2008). *The Icfai University Journal of Banking Law*, Vol. 2, April 2009. Available at SSRN: http://ssrn.com/abstract=1413765

Kash, T. J., & Darling, J. R. (1998). Crisis management: prevention, diagnosis and intervention. *Leadership and Organization Development Journal*, *19*(Issue 4), 179–186. doi:10.1108/01437739810217151

Katayama, M. (2009). Improving recession probability forecasts in the US economy. Working paper, Louisiana State University.

King, T. B., Levin, A. T., & Perli, R. (2007). Financial market perceptions of recession risk. *Working paper, Federal Reserve Board.*

Mihai Yiannaki, S (2008). Regulatory Failure and the Subprime Credit Crunch - The

Nalewaik, J. J. (2007). Estimating probabilities of recession in real time using GDP and GDI. *Working paper, Federal Reserve Board.*

Siomkos, G. J., & Kurzbard, G. (1994). The Hidden Crisis in Product-harm Crisis Management. *European Journal of Marketing*, *28*(Issue 2), 30–41. doi:10.1108/03090569410055265

Stock, J. H., & Watson, M. W. (1993). A procedure for predicting recessions with leading indicators: econometric issues and recent experience. In Stock, J. H., & Watson, M. W. (Eds.), *Business cycles, indicators, and forecasting* (pp. 95–153). Chicago: University of Chicago Press.

Wright, J. H. (2006). The yield curve and predicting recessions. *Working paper, Federal Reserve Board.*

Yiang, Q., Yao, D., Garnett, J. & Muller, K (2010). Using a Trust Inference Model for Flexible and Controlled Information Sharing During Crises. *Journal of Contingencies and Crisis Management,* Volume 18, Issue 4, pages 231–241, December 2010. Wiley. DOI: 10.1111/j.1468-5973.2010.00621.x

KEY TERMS AND DEFINITIONS

Balanced Scorecard: The balanced scorecard in general is a strategic planning and management system that is used widely in any business activity worldwide to attain vision and strategy of the respective organization, improve information flows, monitor organization performance against strategic goals. It was originated by Drs. Robert Kaplan (Harvard Business School) and David Norton as a performance measurement framework that added strategic non-financial performance measures to traditional financial metrics to give managers and executives a more 'balanced' view of organizational performance.

Crisis: 'Negative impact turning point', while from a business perspective, it might be a substantial, unforeseen circumstance that has a higher probability than usual to jeopardize a company's stakeholders, products, services, fiscal or legal situation, or reputation. Both definitions contain an element of urgency that requires immediate decisions and actions from management and employees involved and most possibly from other stakeholders.

Crisis Management: The process of preparing for and counteracting an irregular negative event in order to prevent eventually the organization to fall apart. Crisis management resides in a set of well-coordinated actions to implement thorough control of any potential damage and preserve or re-establish public confidence in the system under crisis.

Entrepreneurship: We refer in the chapter text to entrepreneurship regarding SMEs, as its most obvious form of starting a new business, including its recent association to its terms of social and political forms of entrepreneurship.

Operational Risk: The risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events. We define SMEs Operational Risk as any risk which is not categorized as the risk of loss arising from various types of human or technical error either internally or externally triggered.

Risk Assessment: As a component of risk analysis, it involves identification, evaluation, and estimation of the levels of risks or hazards that have a certain probability to get involved in a situation, their comparison against benchmarks or standards for a particular industry or category of business, and its determination of an acceptable level of risk for the business.

Risk Management: We considered risk management for SMEs, as that particular combination of strategies, policies, procedures, and practices involved in the identification, analysis, assessment, control, and avoidance, minimization, or elimination of unacceptable risks emerging from future events. A company can use risk assumption, risk avoidance, risk transfer, risk retention or any other strategy in risk management. Risk Management is a planned method of dealing with uncertainties to estimate expected and unexpected losses. Risk Management essentially implies identification, measuring, monitoring and controlling risks. Identification of risk is the first and essential step in risk management to assess the extent to which risk could or should be taken and put in place risk mitigation measures. Risk Management is an on-going process as the dynamics change with time. Guidelines on risk management provides broad parameters and each firm may evolve their own systems compatible to their risk management architecture and expertise.

Risk Measurement: Choosing to ignore risk is the riskiest thing a firm can do. One does not measure risk as a stand-alone metric, instead continually analyses the environment to anticipate and avoid situations that may potentially damage the organisation. Some things are thought to be immeasurable only because it is believed that measurement must be some exact value. But the more pragmatic scientific approach to the term measurement is to treat it as quantified uncertainty reduction based on observation (Hubbard, 2010).

Risk Mitigation: We considered risk mitigation or risk reduction the systematic and managed decline in the level of exposure to a risk and/or to the odds of its occurrence.

Sustainable Innovation: The term of sustainable innovation in connection to the risk management of SMEs relates to that particular management approach that leads any new product or service, companies to analyse the benefits, risks and impacts of not only economic factors, but also social and environmental ones.

This work was previously published in SMEs and Open Innovation: Global Cases and Initiatives, edited by Hakikur Rahman and Isabel Ramos, pp. 238-257, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 48
Aircraft Development and Design:
Enhancing Product Safety through Effective Human Factors Engineering Design Solutions

Dujuan B. Sevillian
Large Aircraft Manufacturer, USA

ABSTRACT

Effective Human Factors Engineering (HFE) has provided the aerospace industry with design considerations that promote aviation safety in the development of complex aircraft systems, as well as the operators and maintainers that utilize those systems. HFE is an integral aspect within the systems engineering process. Measuring the effectiveness of Human Systems Integration (HSI) in the research & development stage is critical for the design of new and modified systems. This chapter focuses on how providing effective HFE design solutions enhances product design and system safety. Providing the customer with safe and reliable products augments mission capabilities throughout the product lifecycle.

INTRODUCTION

Multi-modal transportation such as aviation, marine, rail, transit, cycle, pedestrian, and motor vehicles all provide the general public with choices for determining which mode they prefer to utilize. The cost of travel, safety, and transportation reliability are all factors the public uses to determine their travel choices. From a safety perspective, federal agencies have provided the general public with safety statistics for all modes of transportation.

The National Highway Traffic Safety Administration (NHTSA) released data on fatal car crashes in 2008. According to the Fatality Analysis Reporting System (FARS) of the NHTSA, there were

DOI: 10.4018/978-1-4666-1945-6.ch048

more than 34,000 people killed in vehicle crashes in the U.S. (Administration, 2008). The Bureau of Transportation Statistics (BTS)-Transportation Statistics Annual Report-2008 report indicated that there were 4,654 fatalities as a result of pedestrians struck by highway vehicles, and 16,520 passenger car occupant fatalities in 2007 (Statistics, 2008). The public continues to manage their personal risks regarding travel choices and determining which mode of transportation is the safest.

Studies conducted by the National Transportation Safety Board (NTSB), a U.S. independent federal agency that investigates transportation accidents, have shown that commercial aviation transportation is a vital asset to the U.S. economy and is one of the safest modes of transportation. According to former NTSB chairman Marion C. Blakey, "The U.S has long depended on aviation, not just as a means of point-to-point travel, but as the hinge on which the door of commerce depends to stay open". "Literally, aviation keeps America open for business" (Blakey, 2007, p.1). Aviation accidents declined between 2008 and 2009. The NTSB conducted a study that demonstrated a reduction of aviation accidents in 2009—where "The total number of U.S. civil aviation accidents decreased from 1,658 in 2008 to 1,551 in 2009". "Total fatalities also showed a decrease from 566 to 534". "On-demand Part 135 operations reported 47 accidents in 2009, a decrease from 58 in 2008". "The accident rate decreased to 1.63 per 100,000 flight hours in 2009 from 1.81 in 2008". The NTSB have also been able to pinpoint that the "the majority of these fatalities occurred in general aviation and scheduled part 121 operations" (Board, 2010, p.1).

Even though commercial aviation transportation is considered a safer way of traveling, incidents and accidents continue to occur within the industry. From an international aviation perspective, high accident rates in Africa and Indonesia continue to be a critical factor in the general public's perception of air transportation safety. Consequently, federal agencies across the globe have made substantial progress in determining influential factors that cause aviation incidents and accidents. In particular, the goal has been to reduce the rate of occurrences and provide the public with a positive image of aviation safety.

The military and civilian air transportation sectors have developed robust strategies of aviation safety management. Federal government officials, manufacturers, airlines, and independent consultant agencies have implemented aviation safety management methods for the continuous surveillance of safety in flight operations and maintenance. These methods were created to reduce the rate of incidents and accidents through various systematic processes. These systematic processes were developed to manage the 'common threads' in incident and accident causation.

After the crash of ValuJet flight 592 on May 11, 1996, the U.S. implemented the Air Transportation Oversight System (ATOS). The ATOS is an inspection system that was designed to systematically manage safety in flight operations and maintenance in the airline industry. Other reliable methods have all provided the airline industry with structured systematic ways of managing aviation safety. Some of these methods are as follows: Aviation Safety Action Program (ASAP), Maintenance Safety Action Program (MSAP), Dispatch Safety Action Program (DSAP), Safety Management Systems (SMS), and the Aviation Safety Information Analysis and Sharing System (ASIAS). Accident investigation, empirical, and theoretical research continues to provide data useful for determining the causal factors of incidents and accidents.

Aircraft design, weather, and human factors have been a constant trend of incident and accident causation throughout the history of the air transportation system. On February 12, 2009, Colgan Air flight 3407 crashed outside of Buffalo, NY with no survivors. The probable cause of this crash was "the captain's inappropriate response to the activation of the stick shaker, which led to an aerodynamic stall from which the airplane did not recover" (Board, Loss of Control on Approach, Colgan Air, Inc., Operating as Continental Connection Fight 3407, 2010, p.1). On August 27, 2006, a Comair Regional jet crashed in Lexington, KY with only one survivor. The probable cause of this accident was "the flight crewmembers' failure to use available visual cues and aids to identify the airplane's location on the airport surface during taxi and their failure to cross-check and verify that the airplane was on the correct runway before takeoff" (Board, Attempted Takeoff From Wrong Runway; Comair Flight 5191, 2007, p.1). The NTSB investigations revealed that the ValuJet, Colgan Air and Comair crashes were considered to be related to the concept most often referred to as 'Human Factors'. According to the Human Factors Accident Classification System (HFACS) Analysis of Military and Civilian Aviation Accidents: A North American Comparison; "Between 70-80% of aviation accidents are due, at least in part, to human error" (Wiegmann, 2004 p. 2).

Historically, Human Factors related deficiencies in air transportation have influenced incidents and accidents across the global airline industry. In December 2003, a ramp worker lost his life while preparing to push back an aircraft. According to the report, the tractor operator's foot accidently hit the accelerator and "the tug jumped forward striking the radome of the aircraft fatally pinning the driver". The investigation revealed that the ramp agent did not set the parking brake (Matthews, 2004 p.6).

The constant trend of human factors related incidents and accidents in the aviation industry appear to be a parallel theme in other modes of transportation such as the railroad, chemical, highway, and marine industries. On June 11, 1997, a bus collision with pedestrians in Normandy, Missouri occurred. The causal factor in this crash was, "The driver trainee's misapplication of the accelerator resulting in the bus's override of the curb and travel onto the occupied pedestrian platform" (Board, Bus Collision with Pedestrians, 1997). On October 12, 2003, a Northeast Illinois Regional Commuter Railroad train derailed. The probable cause of the derailment was "the locomotive engineer's loss of situational awareness minutes before the derailment because of his preoccupation with certain aspects of train operations that led to his failure to observe and comply with signal indications" (Board, Derailment of Northeast Illinois Regional Commuter Railroad Train, 2003). Even in the design of Unmanned Air Vehicles (UAVs), "the Air Force Scientific Advisory Board (AFSAB) identified the human system interface as the greatest deficiency in current UAV designs" (Williams, 2006 p.1).

The term Human Factors is frequently utilized in the aviation industry. Human Factors can be operationally defined in this chapter as factors influencing the behavior of the human based on several dynamic critical interfaces. Generally speaking, the interfaces that affect human performance include: environmental conditions, software/hardware factors, and interactions with other humans. HFE is the application of human capabilities and limitations as they relate to the design of systems. HSI is the systematic process of integrating the human with the system so that the human has the ability to operate, maintain, and support personnel needs within the applicable environment. HSI is needed to reduce possibilities of human performance related errors and to promote aviation safety. These definitions will be further explained throughout the chapter. Additionally, HFE and HSIs relationship to overall aircraft systems design will be discussed.

Reducing the possibilities of human error can be a rigorous task. The industry has developed several methods to aid in the reduction of human performance based errors. Some methods such as: HFACS, Maintenance Error Decision Aid (MEDA), and the Ramp Error Decision Aid (REDA) are internationally accepted methods designed for human error reduction. Essentially, these methods are utilized to help manage the risk of human error and to better prepare the human for the tasks and environmental conditions they will experience.

Several years ago, the term 'pilot error' was commonly used by air safety investigators to describe an individual(s) flying an aircraft that was involved in an aircraft incident or accident. It is paramount to understand that 'pilot error' is not synonymous with the term human factors. The refined definition of human factors requires a complete understanding of the 'human element' as a system interfacing with a series of systems. The process of integrating the 'human element' within a system is not a parallel process within the systems engineering discipline. Therefore, understanding the dynamics of the human system interface continues to be a challenging and critical aspect in the development and design of complex aircraft systems. Nevertheless, one might ask the question, 'How is a system designed without human error?' It is impossible to design a system without the possibilities of human error. In fact, as long as there is human interface with a system, there will be the potential for errors. As engineers, investigators, scientists, and specialists, we must determine the impact of the system design on human performance, while minimizing the potential risk associated with the hazards.

The goal of this chapter is to provide the reader with a clear understanding of the need for human factors considerations at the design stage, and to examine the factors that influence safety throughout the product lifecycle. This chapter will provide case scenarios related to human factors and system design; to further demonstrate the relationship between the human and its interfaces. Furthermore, there will be a discussion of human factors issues. Finally, solutions, industry recommendations, and future research directions will be provided. An outline of topics that will be covered throughout the chapter are as follows:

- General understanding of system engineering and design architecture
- Human integration vs. system design
    - Cognitive interfaces on system design
    - Physical interfaces on system design
- Enhancing product integrity through promoting effective system safety management
    - Hazard Identification
    - Hazard Analysis
    - Hazard Mitigation
    - Hazard Monitoring

BACKGROUND

Since the Wright Brothers flight on December 17, 1903, the aviation industry has learned from, scientists, pilots, engineers, and other aviation professionals, the need to develop safe technologically advanced aircraft systems. Since World War II, the design of aircraft systems, and the apparent human factors influence on these systems has increased the ability for research and development in the human performance domain.

The aerospace industry has learned how to design aircraft systems that are reliable, safe, cost effective, and efficient for the crewmembers that utilize those systems. Recent developments in aircraft design and test, such as the Boeing 787 Dreamliner and the Boeing 747 Large Cargo Freighter (LCF)-Dreamlifter further re-iterates the continuous strategies and perspectives of safety through constant innovative new technologies. However, the design of complex aircraft systems can be a challenging effort. An abundance of engineers are required to produce quality products that are safe, human compatible, reliable, and meet the customer mission requirements. As the complication and interdependence in systems grow, the certification and testing of those systems continues to be a crucial factor in aircraft system development (Abbott et al., 1999).

It is inevitable that when engineers design a system, there will be some type of human interface with the product. Whether the system is a manned or an unmanned crew station, human system interface continues to be a persistent issue when designing any aircraft system. Software and hardware technological advances have provided the industry with the opportunity to design systems that have components that are interconnected. However, these systems have the potential to fail (Abbott et al., 1999). These failures can be considered systematic failures, those related to the aircraft system functions and human system interface failures, and those related to the integration of the human with the system.

Failures related to the system reflect deficiencies in the way a system was executed. These insufficiencies are due to lack of training, personnel selection, operating policies, maintenance and organizational support issues. Human system interface failures are design errors which include hardware, software, environment, and training factors. These design errors are the result of insufficient integration of the human with the system. In particular, Graphical User Interface (GUI), and other ergonomic equipment design related issues are some of the factors that influence the user interface (McSweeney and J.Pray, 2009).

It is important to understand that having a technologically advanced system does not constitute a safe and user-compatible product. The human utilizing the system has to be able to effectively operate and maintain the system. If the operators and maintainers cannot effectively utilize the system, then the potential of human error exists and the design becomes incompatible for the user. In order to reduce the possibilities of design human incompatibility that could lead to induced human error, an understanding of the system design architecture becomes a crucial aspect at the design inception stage. During this stage, both systems engineers and design engineers work together to understand the system fundamentals.

A system is defined as an interactive integrated composition of architecture that work together to achieve a specific mission. According to Chapanis (1996), there are three views of a system: the physical view, functional view, and the operational view. The physical view is based on the population of people that the system will contain. Next, the functional view focuses on how the system should work from an operational perspective. Last, the operational view is based on how the users interface with the system. Systems engineering is the iterative integrated process of managing the development of complex systems throughout the product design lifecycle. Without systems engineering, it would be difficult to manage the development of any aircraft system because systems engineering provides the framework for managing the design.

MAIN FOCUS OF THE CHAPTER

"The Iterative Process of Aircraft Development & Design"

When a customer solicits participation in an aircraft design contract, they want a contractor that can provide customer service that will support their mission needs. Whether these customer mission needs are to fly passengers from point A to B, or to fly a secret mission, it is paramount when designing any aircraft system to implement a 'systems thinking' tactic to design. A 'systems thinking' approach to product design is a disciplined scheme to problem solving. Typically, this process can be defined as how certain systems influence the aircraft as a whole. The approach also focuses on how the human interfaces with a system and its components. Utilization of 'systems thinking' should occur when determining the purpose of the system, functional capabilities, and operational capabilities. Finally, this approach requires knowledge of five general integrated areas related to systems engineering. The five taxonomies are as follows:

- Product definition
- System and mission requirements
- Functional Analysis/Functional Allocation
- System Design
- System Test

When designing an aircraft system, it is important to understand the conceptual design stage, which is most commonly referred to as the product definition. At this stage, the forecasting of engineering strategies, business planning development, and proposal planning are considered and executed. Reputable competition strategies are a key asset within the product definition stage. While competition strategies are key to the development and design of aircraft systems, system engineers create program plans to effectively collaborate and manage the process of design. The HSI Program Plan, Human Engineering Program Plan (HEPP), System Safety Program Plan (SSPP), Reliability/Maintainability Program Plan, and System Engineering Management Plan (SEMP) are just some of the program plans that should be developed at this stage. These collaborative program plan approaches to engineering design should be reviewed with the customer. These plans are needed so that the customer has a clear understanding of how the engineering process will be managed.

The contractor Statement of Work (SOW) provides the foundation for understanding baseline level customer requirements. From a system engineering perspective, a requirement defines a necessary element or characteristic that must be satisfied in order to achieve operational capabilities. A requirement should be further emphasized in order to determine how it applies to the system and the environment.

System and mission requirements are essential components within the systems engineering process. System requirements describe 'what' a system shall do from a hardware and software perspective. Mission requirements are performance specifications and they relate to 'how' the system shall perform in a particular environment. For example, a system requirement may be, 'the software shall be compatible for the display user interfaces'. A mission requirement may be, 'the satellite shall be capable of differentiating between various types of surface matter'. The impact of these requirements must be further explained and reviewed by the customer and contractor. Trade studies help systems and design engineers understand how the system and mission requirements impact the user operational and environmental characteristics. It is imperative that these requirements are thoroughly examined in order to ensure consistency in the application of the requirements in system design; this is needed to ensure the requirements are attainable, verifiable and valid.

The functionality of a system drives certain requirements. Whether the system is electrical, hydro-electrical, mechanical, or pneumatic, a complete understanding of how the system functions relative to the requirements assists the designer with developing the system. Furthermore, the function of the system assists the designer with how the system will perform in an operational environment. Examples of system functionality in an operational environment are the aircraft phases of flight. Typically, the aircraft phases of flight are as follows: pushback, taxi, takeoff, climb, cruise, decent, and landing.

The function of an Aircraft Communications Addressing Reporting System (ACARS) is to provide the aircrew with a digital data link utilized as an interface between dispatch and crewmembers. Some of the characteristics of the ACARS system are the capability to send information such as aircraft system performance data characteristics, generation of weather reports, flight plan management, and maintenance message communications. The data generated from the ACARS ease communication with airline System Operational Control Centers (SOCCs). SOCCs are the communication command centers at the airline base of operations that ensure the aircraft and crew members are operating adequately throughout the phases of flight. To ensure that the crew is operating satisfactorily, maintenance control, dispatch, and other related airline functions are communicated through the use of the ACARS. The ACARS system is utilized in all phases of flight.

Another system that is utilized in flight as a power generator is the Auxiliary Power Unit

(APU). The function of the APU is to provide the aircraft with an additional power source. This power may be used for transferring air to the cabin for air-conditioning purposes or electrical power distribution. The APU also assists the crew in an emergency situation (engine loss, or failed power generator) by providing the crew with an alternative means of power. Moreover, the function of the system must be evaluated so that the system engineer understands how to meet customer requirements and the impact of system operations.

Typically, after the functions of the system have been defined and analyzed, the allocations of the functions are evaluated. Functional allocation is "a procedure, for assigning each system function, action, and decision to hardware, software, human operators, or maintainers or some combination of them" (Chapanis, 1996, p.101). Functional allocation is an important concept because it assists the designer with an understanding of the viability of technology related to software and hardware interfaces. Additionally, the software and hardware interfaces should be reviewed by HFE to determine the practicability in meeting the requirements. The data evaluated at this stage of the process are previous experiences with systems, technologically advanced performance capabilities or mechanical systems and software systems related to human limitations. Functional flow diagrams help facilitate and understanding of the functional allocation process (Chapanis, 1996).

The system and mission requirements are reviewed by the contractor and customer. These requirements are derived and further delineated through a series of customer reviews. These reviews help the customer and contractor understand the feasibility of the design for future product developments. Depending on the mission, the customer may choose how they want certain requirements implemented in particular stages of the system design; this is normally based on current and future design strategies.

An Operational Readiness Review (ORR) determines the feasibility of proceeding to the concept exploration. The inputs to this review are the operational need analysis and the operational concept analysis (Chapanis, 1996). The contractor develops the concept document to determine if customer needs are satisfied.

A System Requirements Review (SRR) is then conducted with the customer to determine how the requirements will be utilized to design the system. Depending on the program, there may be a series of SRRs developed to further delineate the requirements through incremental stages of the contract. The SRR also examines the various methods of compliance to the requirements and the methods the contractor has chosen to perform analyses to meet the requirements.

Normally, the System Design Review (SDR) is conducted after the system specifications have been reviewed. This review is required to determine if the integrated system will meet the requirements (Chapanis, 1996). The human factors requirements that are ordinarily reviewed (as they relate to the overall system) are as follows: Human Computer Interaction (HCI), ergonomics, safety, operations, maintenance, personnel requirements, and training.

During the Preliminary Design Review (PDR), the customer, program management, and various engineers are assembled to review the current design philosophy. The PDR includes the approach to various design configurations such as hardware component specification, software specifications, and detailed requirements for future test evaluations. Specialty engineers provide support to the PDR approach and they include: Reliability Maintainability & Testability (RM&T), Electrical, HFE, and system safety engineering.

After the PDR, the Critical Design Review (CDR) is coordinated. This review is the final review before the production stage of the system. The purpose of this review is to compile all of the detailed engineering data related to software and hardware configurations. At this stage, engineering drawings, mockups, HFE, safety, reliability & maintainability analyses are reviewed. Engineering data is provided to the logistics team in order to update or develop new training methods and procedures to support crewmembers in the operational environment.

Last, the Test Readiness Review (TRR) is conducted to review the design integration as it relates to the capability of the system to operate effectively in the operational environment. The design is tested in regards to the interface between the human and the system to ensure user compatibility and safety; various test plans are developed to test system effectiveness. Test reviews are performed to ensure that the contractor has met all of the customer requirements and that the system is suitable for the operational environment.

"The Human Integration Strategy"

Historically, the systems engineering process has been portrayed as a complex integrated approach to product design. The process has worked for several years in the development of aircraft systems. Each stage of the system engineering process conveys a common theme. This theme is related to the continual need to ensure 'system compatibility' so that the system can function in the operational environment. It is pertinent for design engineers to develop systems utilizing their knowledge of engineering design. However, a more important aspect is the design of the human into the overall system design.

Most of the literature related to system design covers a wealth of information regarding overall system design effectiveness and operational support. However, it is argued that there is a consistent issue with the value, understanding, and the application of implementing a HSI approach to engineering design. In addition, HSI is often considered an 'excess cost agent' to program management and to the customer. How should a system be designed to encompass program and customer requirements, and cost dynamics, while also providing the customer with a safe and effective integration of human system interfaces? This is the primary question that most human systems integration engineers, human factors engineers and specialists ask themselves when participating in any program.

The U.S Air Force, Federal Aviation Administration (FAA), U.K. Ministry of Defense and the Canadian armed forces have all incorporated HSI within their Human Factors commercial and defense programs. HSI is the comprehensive management strategy for effectively implementing the human with the system, in order to fully support the human interfaces considering various conditions.

This strategy provides, tools, techniques, to help facilitate a more robust effort for human integration within systems. Previous approaches only focused on hardware and software integration issues. This strategy focuses on the end user and the relative operational dynamics (Reinach, 2007). Qualitative and quantitative realistic approaches should be utilized to evaluate the risk and to measure the need to utilize a HSI approach in the program development stage. In particular, trade studies should be used to provide evidence to the program regarding the need to implement a HSI program.

Customer program management, contractor program management, design engineers, and system engineers should all interface with HSI. According to the Defense Acquisition Guidebook (DAG), there are seven domains applicable to supporting the integration of the human with the system. The seven domains are:

- Manpower
- Personnel
- Training
- Environment, Safety, and Occupational Health
- Survivability
- Habitability
- HFE

"Manpower can be a major determinant of program cost and affordability" (Guidebook, 2010, p.2). Manpower focuses on operations and maintenance tasks and conditions in the operational environment that are associated with workload. Moreover, the manpower perspective focuses on the need to support military and civilian manpower. Manpower considerations are important to the customer. Program management should consider developing case studies and engineering analyses to optimize manpower and reduce costs to affordable levels. Effectively utilizing manpower considerations in system design can reduce the cost associated with design and operations. One factor that reduces the cost of manpower is an understanding of user tasks as they relate to the system functional allocation processes. If the engineer understands the user tasks, then only relevant tasks will need to be evaluated. These tasks are unique to the system design, and are related to cognitive, physical, environmental, and other related conditions. Evaluating these parameters helps the customer determine proper manpower considerations for their program.

"Our uniqueness as individuals--the way we were raised and educated, our work experiences, and genetic compositions--affect the way we perceive the world and act upon it" (Strauch, 2005, p. 47). Personnel considerations are developed based on the Knowledge, Skills, and Abilities attributes (KSAs) of individuals. The KSA attributes are related to the cognitive and physical demands of the system versus the demands for personnel needed to perform the jobs (Guidebook, 2010). These factors are used to determine the experience levels that operators and maintainers need to perform their job task. KSAs help engineers understand human capabilities and limitations as they relate to system design and personnel qualities. Lastly, personnel selection helps engineers design tasks properly in order to reduce the possibilities of human error.

"Training is the learning process by which personnel individually or collectively acquire or enhance pre-determined job relevant KSAs by developing their cognitive, physical, sensory, and team dynamic abilities" (Guidebook, 2010, p1). When the customer and program management have completed the CDR and the customer is satisfied with the design, training factors need to be discussed with the customer as they relate to the new/modified design. These factors include simulation training, On-the-Job-Training (OJT), or Computer Based Training (CBT). After these factors have been addressed, it is paramount that when personnel are trained, they understand the new or modified system functions and their components. Ongoing training should assist personnel with the new/modified operating procedures and new or modified system capabilities (Strauch, 2005). Essentially, documents such as Aircraft Maintenance Manuals (AMMs), Flight Crew Operations Manuals (FCOMs), and Technical Orders (T.O.s) are just a few of the documents that may need to be updated as a result of the new/modified system design. These updates are needed to ensure that the new/modified equipment design aligns with the nucleus of the training program curriculum.

Environment, Safety, and Occupational Health (ESOH), are essential components to the HSI process. Understanding the factors that influence the surroundings of the human when operating and maintaining a system can provide designers with an understanding of how to accommodate the user with the system. The environment includes "the conditions in and around the system and the operational context within which the system can be operated and supported" (Guidebook, 2010 p.1). In particular, certain environmental issues should be evaluated by engineering. Some of these issues include acoustical noise, vibration, and electric shock. Engineers need to determine the safety impact on the human and the mission as a result of environmental conditions. Safety related issues such as workspace design, Personal Protective Equipment (PPE), hazardous waste, and radiation need to be evaluated to determine the impact on the human and the mission. Warnings and cautions that are depicted on systems or within workspace areas should be clear and concise. These placards should provide the user with enough information to understand the warnings or cautions and provide meaningful information for the user to make the most practical decisions. Warnings and cautions should be effectively communicated in training manuals, test procedures, and OJT. While understanding the impact of safety on the human and mission is practical, it is also important to realize the effects of unsafe designs on the human in a particular environment. The occupational and health perspective is incorporated within the design to explain the potential risk of the design on the human body that could lead to musculoskeletal issues, injury, illness, or disability.

Providing the program and customer with an assessment of the occupational hazards as well as a, relative cost benefit analysis helps the customer with evaluating the risk. This evaluation can be used to determine if a design is suitable for the user population. Data presented by occupational safety specialists at the CDR provide the customer with an early indication of the hazards and the applicable mitigation processes.

The survivability domain is most often related to the military or defense environment. Survivability factors are related to design features that reduce the risk of man-made hostile environmental conditions. Survivability design and operational techniques should be outlined in the program requirements. All requirements should be further defined to provide the engineer with an understanding of the type of environment the human will be susceptible to versus the system design. Certain clothing and survivability gear may be utilized to prevent the possibilities of death or the abortion of a mission. An example of survivability gear is the Nuclear, Biological, and Chemical (NBC) gear. This gear should be designed so that the user can perform the mission operations while minimizing the potential risk of injury or fatigue. In particular, some factors to consider when designing NBC gear are egress capabilities, and extreme environments. It is important that the crewmember has the ability to exit the aircraft immediately and survive conditions on the land or sea. HFE assists with making decisions that may affect the design of the user interfaces.

Habitability focuses on the living and working conditions for the user population, in order to effectively accomplish a mission. Depending on the mission, the customer may require only certain aspects of habitability. Some characteristics of habitability are as follows: ambient lighting, food servicing, lavatory service, and medical support. HFE and other related specialists should determine the effect of the design on the user, and develop methods for evaluation to optimize user performance in the habitability environment.

HFE is the application of human capabilities and limitations as they relate to the design of systems. HFE should be implemented within a program to reduce manpower and provide effective crew training. Previously stated, when the system functional allocation is performed, the user interfaces should be identified and evaluated by HFE. HFE should be applied during the development stage of equipment design; this application provides the customer with an understanding of the human machine interfaces. Any related tools (modeling, prototyping, simulation) or techniques (Crew Station Working Groups (CSWGs), workload analysis, strategic processes) should be outlined in the HEPP to address how the user interfaces will be managed throughout the product design stages. Generally speaking, user interfaces include: display design, GUI, maintainability, training systems development, training manual development, workload analysis, and Ground Service Equipment (GSE) support. All interfaces should be reviewed with the customer and program management to understand the cost, and design impacts. A meticulous technology strategy review should be utilized when determining the best method to reduce cost associated with modeling and simulation because these cost drivers could affect the program budget. Understanding the customer requirements for HFE prior to development can reduce the cost of design by eliminating extraneous methods of analysis and evaluation. HFE should be cost effective for the program while enhancing user interface capabilities. If a HSI program is developed, a complete understanding of the HSI domains helps with system development and design, thus providing the customer and the program with data for managing the factors influencing system design.

HFE: "Indispensable Solutions in Design"

So far, the issues relative to transportation safety, system design, and the interfaces related to HSI have been acknowledged. Within the confines of HSI, the remainder of this chapter will focus primarily on the HFE domain of HSI. Also included will be a discussion regarding product safety as it relates to managing system design.

HFE can be separated in to two distinct integrated categories: cognitive engineering, and hardware/human physical ergonomics engineering. To understand the context of cognitive engineering, one must first understand the definition of cognition. Cognition is the study of how the human thinks, perceives and remembers information in a certain environment. Cognitive engineering is a multidisciplinary field focused on the design and evaluation of systems technology as it relates to human interfaces and cognitive processes (Gersh, 2005). Since the Three-Mile Island nuclear power plant incident in 1979, a focus on how the human interfaces with workplace design and socio-technical systems continues to be evaluated by scientists and engineers. In addition, human interface with system controls and displays has been on the vanguard of cognitive scientific research. Cognitive engineering may focus on HCI as it relates to GUI or other cognitive demands on the human in a particular environment. Cognitive engineering also focuses on workload and human performance in specific environments. Hardware/physical ergonomics engineering is the application of human dimensions, physics and biomechanics on human system design integration in a specific environment.

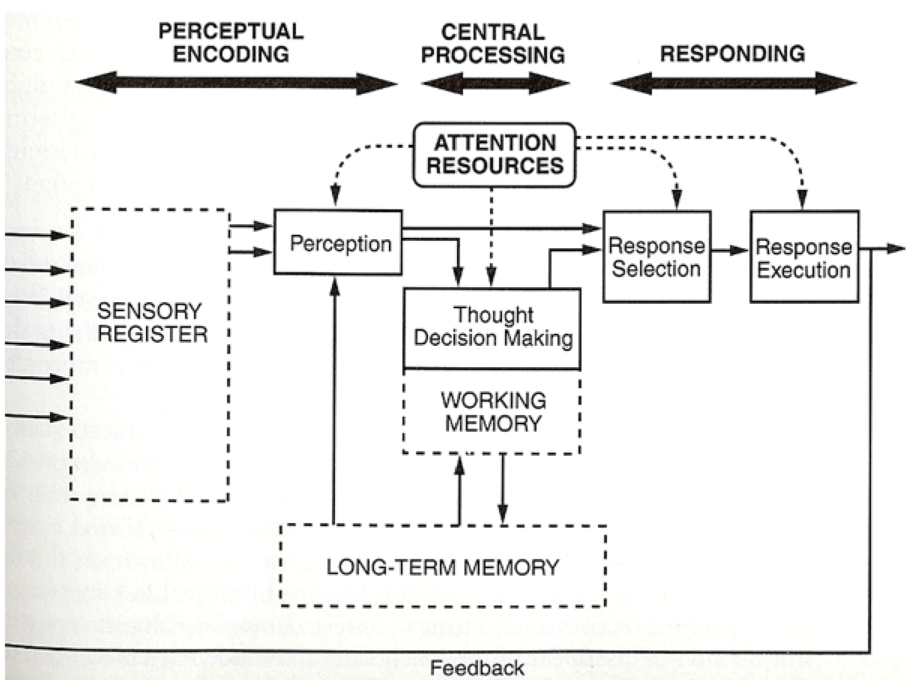
From a cognitive perspective, when functional allocation decisions are considered, HFE should assist the program with making those decisions. It is imperative that once the functional allocation decisions have been made, that HFE outlines how those functions interface with the existing requirements. HFE should provide the program with an overview of how the system functional allocation may affect the human capabilities and limitations. For example, if the requirement is for the human to be able to control system functions, then HFE has to determine the impact of those functions and how they influence the ability of the human to perform certain tasks. These tasks should be based on data obtained from a system functional analysis. The data should be utilized to determine if the functions of the system interface are compatible with the user prospective environments.

After determining the impact on the user and system functional capabilities, the HFE should perform a task analysis. A task analysis is utilized to identify tasks and sub-tasks related to the user environment, and comparing those tasks to the operator and maintainer cognitive processes in order to achieve a goal. Essentially, the task analysis is utilized to: reduce potential hazards and errors, increase human reliability in system interface, evaluating user-system processes, and analyzing personnel needed to perform certain tasks such as operator related tasks or maintainer related tasks. The data generated from task analyses provides information to logistics for developing training documents and other related OJT. The task analysis should describe the human system interactions as they relate to human capabilities and limitations. The results from a task analysis drive new personnel and manpower requirements; these requirements may indicate a need for additional crewmembers or a reduction in crewmembers. A task analysis would be practical if the HFE is evaluating HCI on an operator aircraft display at a particular crew station. The HFE should evaluate the ability of the user to process information on the display screen while performing a task. Whether the task is evaluating the integration of overlays on a display or determining the color and text size, it is important to evaluate how the human processes this information. In order to perform the task analysis, the HFE should have a clear understanding of the Human Information Processing System (HIPS) theory.

According to Wickens (1998), there are three stages in the HIPS theory. The perceptual stage focuses on how the human perceives stimuli and compares it to memory and knowledge to give it meaning. The cognitive stage focuses on how we relate new information to memory or goals to solve problems or make inferences. The action stage is the final stage of the HIPS theory that focuses on how the central processing unit (brain) responds and coordinates utilizing certain motor signals. Once the HFE understands the HIPS theory, the HFE should utilize the Information Processing Model (IPM) to examine all human system interactions to clearly understand how the human processes information related to the system design in a particular environment. The IPM is a model that portrays how the human interprets a stimulus and responds to certain system inputs in a particular environment. Figure 1 illustrates the IPM; this is a generic model used for human information processing.

Hardware/human physical ergonomics engineering focuses on human anthropometry as it relates to the system interface. Generally speaking, HFE integration should support at least the 5th percentile (typically who are female) and at most the 95th percentile (typically who are male) as the critical human dimensions. Critical human dimensions such as: stature, weight, chest breadth and depth are utilized when integrating the human with the system. Designing systems for the 5th and 95th percentiles provides a baseline approach for accommodation of the user population. However, depending on the user population (Army, Navy, Air-Force, Civilian), certain case scenarios

Figure 1. Information Processing Model (Wickens, 1998)



may need to be further evaluated to ensure user-system accommodation. These case scenarios are related to certain body dimensions that may be on the outside boundary of the targeted user population for a particular environment such as chest breadth and stature. Figure 2 illustrates the 5th and 95th percentiles and Figure 3 illustrates critical body dimensions.

Locations of controls and displays, workspace design, and system maintainability are just some of the general areas that must be analyzed by HFE to determine the best method to optimize the user interface. HFE industry standards assist HFE with determining the best method for optimizing usability.

The Department of Defense (DOD) and the FAA publish numerous design standards, and design approaches to provide HFE best practices for accommodating the user with the system. If utilized early in the engineering design process, these standards can reduce the cost associated with the development and design of hardware and software systems. These standards and approaches should be outlined in the HEPP to provide the

Figure 2. CATIA V5 5th and 95th percentile Manikins (National Institute of Aviation Research (NIAR) Laboratories)

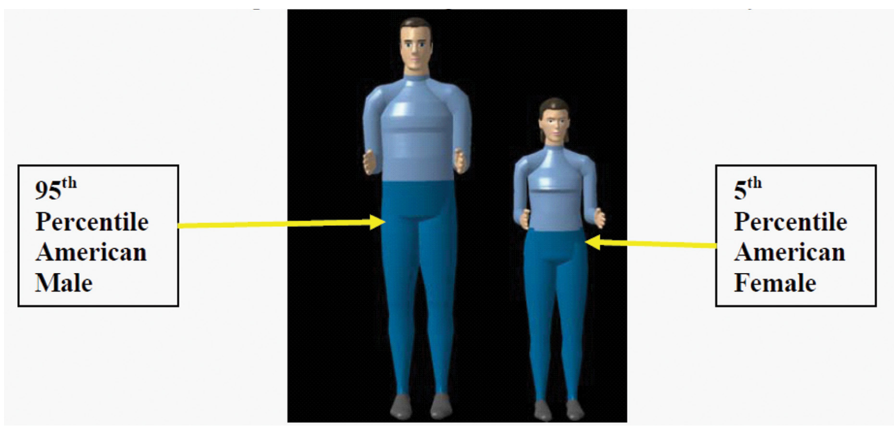
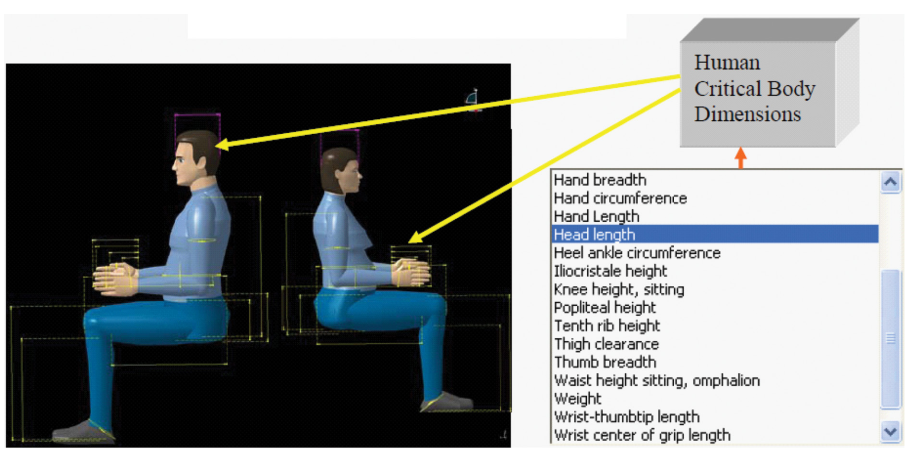


Figure 3. CATIA V5 Manikins critical body dimensions (National Institute of Aviation Research (NIAR) Laboratories)



program and customer with and understanding of how the standards and approaches relate to the system requirements. Furthermore, these standards and approaches provide objective data that can be utilized to substantiate issues related to engineering drawings, computer models and the overall systems integration.

MIL-STD 1472 (DOD Human Engineering Design Criteria Standard) is one of the most common utilized standards by HFE for hardware design. Many of the supplemental HFE industry standards are derived from MIL-STD 1472. From a hardware perspective, this standard provides an abundance of data related to the design of certain hardware such as: knobs, dials, pushbuttons, joysticks, pedals, toggle switches, and commonality in systems, just to name a few. The standard also provides information regarding safety, human visual fields, system color coding, and lighting. MIL-HDBK 759 (Human Engineering Design Guidelines) is utilized for determining various anthropometric variables that can be used for the design of systems. Variables such as arm length and chest breadth assist the design engineer with understanding the impact of human dimensions on system spatial integration.

HFE often utilizes design approach documents to help facilitate robust methods for human design integration. Two of the most common design approach documents are the Human Engineering Design Approach Documents-Operator and Maintainer (HEDAD-O/M). Both documents outline HFE methods to assist the HFE with understanding how to evaluate the human interface with the equipment. These documents provide information on how to meet HFE requirements and HFE design criteria. The approaches should be communicated with the customer during SRR and PDR. Communicating these approaches gives the customer a better understanding of how the operator and maintainer systems will be evaluated.

MIL-STD 1472, MIL-HDBK 759, and other applicable standards should be utilized when creating computerized engineering models. These standards provide the HFE with data to support his/her modeling activities. One of the most common systems utilized by HFE for hardware human integration is the Computer Aided Three-Dimensional Interactive Application (CATIA). CATIA is an iterative process application, utilized to develop human engineering models that predict human performance in system design. Rather than having physical prototypes, human engineering models reduce the cost of human factors analysis by providing the customer with an early indication of possible human factors related integration issues prior to CDR. CATIA software provides the HFE with various workbenches to support HFE analysis. Two of the most common workbenches related to HFE analysis are ergonomics and kinematics. These workbenches aid the HFE with developing digital human static models (manikins) and motion activated models that support maintainability and human crew station compatibility.

Digital human static models are utilized to provide multiple case scenarios related to maintainability and crew station compatibility. These models can be manipulated to show human performance issues in system design. The motion activated models provide the customer with an analysis that conveys the relationship between engineering models (system models) and how they are put in to motion (kinematics) relative to the manikin simulated motion. A Digital Maintainability Analysis (DMA) via CATIA would be practical if there is a need to analyze the human access abilities for Removal and Replacement (R&R) of Line Replaceable Units (LRUs). It is important to ensure that the user can access, inspect, and opt for the R&R of LRUs. This analysis is also needed to meet applicable maintainability requirements. Depending on how long it takes maintenance personnel to R&R, access, or inspect a LRU, this analysis could determine the need to allocate certain personnel or manpower for the task. A tooling analysis provides an evaluation of the types of tools needed to access the equipment and the ability to R&R the system and its components; virtual tools (wrenches, screwdrivers, hammers, etc.) are utilized to determine system accessibility.

The force required to remove a LRU is computed utilizing a Push-Pull Analysis; this helps the HFE understand the effort needed to remove an LRU. A carry analysis is used to ensure the user can carry an LRU with respect to system weight and system lifting characteristics. Variability in system weight should be outlined in the system description document or other applicable equipment detail documents. It is imperative that a thorough evaluation of the effects of system weight on human performance is analyzed. User population and system weight characteristics should be identified and defined prior to PDR to ensure these factors are addressed. Depending on the type of environment, cold weather gear may need to be evaluated to determine user compatibility in maintenance operations. Arctic gloves/mittens may be tailored for a certain user population by simulating certain measurements to the hand. Applying specific measurements to the hand utilizing CATIA helps accommodate the design of arctic gloves/mittens.

A Crew Station Compatibility Analysis (CSCA) via CATIA is utilized to ensure crewmembers are able to operate the systems effectively and that they possess the ability to work in their environment. A CSCA should be utilized when determining the user's ability for grasping controls, reaching display screens, or any other applicable hardware at the crew station. For example, if an automated warning, caution, or advisory display message requires crew input to retrieve information regarding a display message, it is important that crew's ability to access the information is optimized. In other words, this means that the display needs to be designed so that the user has access to the display and has the ability to operate the controls effectively. Hand size dimensions should be evaluated with respect to control system design to ensure that the human has the ability to grasp or press certain buttons.

Line of sight and visual field evaluations are utilized to determine human system visual compatibility. Certain equipment such as Helmet Mounted Displays (HMD) may be evaluated to determine operator spatial orientation and the ability to recognize objects in the operator Field of View (FOV). Essentially, the HMD analysis should be evaluated to ensure the user's ability to effectively perform a particular mission. Additionally, the customer mission requirements may dictate a need to accommodate a certain user population in regards to the design of the HMD.

A human posture analysis should be utilized to determine the need for the operator to maintain a certain posture while operating a system at a crew station. In regards to pneumatic systems, such as crew oxygen, it is important to evaluate the crew's ability to reach oxygen masks while maintaining a certain posture. Furthermore, variability in seat designs such as the ability of the seat to be elevated or lowered is critical to understanding how the seat design affects the user's ability to reach the oxygen mask. In particular, the stature of the human could affect the ability to operate certain controls at the crew station. The analysis should provide the designer with information to develop or modify a seat configuration so that it is compatible for the user.

Table 1 illustrates the types of analysis that HFE should conduct utilizing CATIA in order to

Table 1. CATIA/human system interface design considerations

DMA	CSCA
Human/LRU Access Ability	CSCA Display Reach Analysis
Tooling Analysis	CSCA Controls Grasp Analysis
Push-Pull Analysis	Visual Field Envelope Analysis
Carry Analysis	Human Posture Analysis
Cold Weather Gear Analysis	HMD Analysis

measure the efficiency of hardware human system interfaces related to human performance.

CATIA models should be utilized to communicate with the program and the customer during the PDR and CDR. Providing these analyses assists the program with understanding the impact of HFE on the operator and maintainer interfaces. Lastly, CATIA models help facilitate the integration of system design, while providing assistance with determining the safety impact on system design.

Product Safety: 'A Management Perspective'

Thus far, various meaningful methods for accommodating the user from a cognitive and hardware engineering perspective have been mentioned. These methods are paramount for evaluating user interfaces. However, the need to evaluate user and system compatibility are not the only methods for ensuring product user efficiency.

Product safety management is the ability to manage safety related risks by providing reliable safety analysis suitable for the operational environment. Product safety management is imperative when certifying any aircraft or system. Safety risks are evaluated to determine if an aircraft and system can meet certification requirements to be certified. Various methods of reducing the risks associated with system hazards prior to system design have provided the industry with the ability to engineer-out or to provide other solutions to mitigate certain hazards. These hazards could have a direct influence on the aircraft, system/subsystem and the user. Hazards associated with the aircraft, or system/subsystems could affect mission capabilities, and hazards associated with the system/subsystem user could affect personnel life and work environments.

The most common way of reducing risks associated with a system/subsystem and the user is through the application of system safety engineering. System safety engineering provides the program and the customer with an early identification of risks. A risk is defined as the probability and severity associated with a particular hazard. System safety engineering identifies latent and active failures to reduce the potential of a hazard. Latent failures are related to system/subsystem deficiencies which may be undetected. An active failure is the result of failing to detect system/subsystem deficiencies in which case, the possibilities of an incident or accident are likely. Although latent and active failures occur in system and subsystems, the result of the failure could have an effect on the human. Thorough analysis techniques should be utilized to disclose the risk associated system and subsystem failures and their possible effects on the human. Through the use of fault tree analysis, Failure Modes and Effects Analysis (FMEA), and other related system safety analysis techniques and philosophies, the program and customer have a realistic view of potential risks on the system, subsystem and the human.

During the product definition phase, a SSPP should be utilized to outline the system safety organization plan on managing safety related program risks. The customer should provide the safety requirements to the contractor so that the contractor has an understanding of how the safety requirements may affect the system design. The system safety engineer should determine the best methods to meet those requirements. During the PDR, the system safety engineer provides a preliminary evaluation of the system/subsystem risks during certain phases of flight; risks associated with system human integration are also identified. It is essential that the system safety engineer analyzes the risk from the system component and aircraft level, in order to understand the impact of the risk for the proposed design configuration. From a military perspective, MIL-STD 882 (DOD Standard Practice for System Safety) is used to determine the levels of risk associated with certain designs and operations. ARP-4761 (Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne

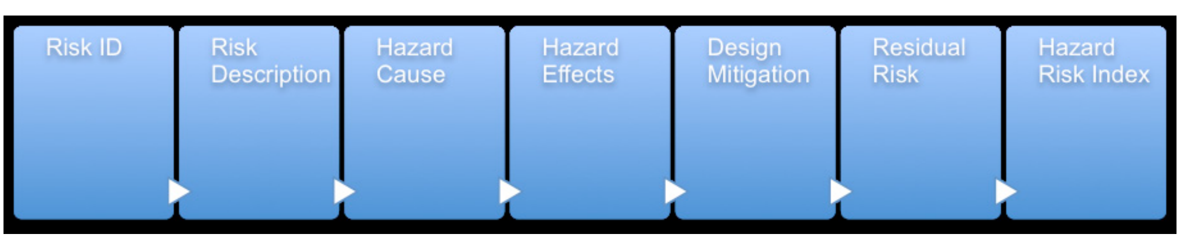
Systems and Equipment) is normally utilized for commercial aircraft or systems. The techniques utilized for identifying risks may be different, but the philosophy of risk identification and mitigation should be transparent and support customer and program goals.

Classically, a preliminary risk assessment starts with the identification and description of the risk. This provides the program and customer with an understanding of the issue. For example, the risk description may be characterized as: 'Possible insufficient installation of LRUs' (Risk Identification #1). All risk descriptions in the preliminary risk assessment should be related to the proposed design configuration and must be clear and concise. The system safety engineer must have a clear understanding of the system description and the system function in order to proceed to the next step of the preliminary risk assessment. The next step of the assessment is to identify the causal factors and the effects of the risk. In this case, the causal factors could be related to the design of the LRUs and the integration in to the aircraft. The effect may be the LRU falling off an equipment support rack, leading to maintainer injury. As a result of the LRU falling off the rack and depending on the LRU location on the aircraft, the potential for inhibiting egress operations for the operator and passengers may exist. Next, the design mitigation strategy and any residual risks are identified. In this case, the design mitigation may be to re-design or relocate the LRUs to ensure the LRU installation configuration is optimized. To ensure the design is optimized, engineering drawings and CATIA models will need to be modified. Customer and program management must understand the engineering change and the potential effects on the mission. Any residual risks (risks left after design mitigation) should be identified and communicated with the customer at PDR. Finally, a preliminary hazard risk index should be provided to identify the probability and severity associated with the hazard. Figure 4 illustrates a typical preliminary risk assessment table outline.

All new or modified risks should be identified as the design develops. As the design matures, certain risks may need to be further communicated with the contractor, program management and the customer prior to CDR. For example, certain risks that could affect mission capabilities or relatively high risks such as radiation hazards should be effectively communicated. It is the responsibility of program management to adhere to recommendations provided by system safety engineering and incorporate certain design features to reduce the potential risk of a hazard. As the program transitions in to test and operations, a thorough evaluation of the hazards should be communicated through a series of safety assessment reports. The reports generated should be inclusive and provide conclusive evidence for the reporting of potential risks. Any analysis relative to potential system level, aircraft level or personnel hazards should be documented and monitored through a tracking system. It is up to the contractor to effectively monitor and provide updates to the customer regarding new or modified hazards.

The overall goal of safety management is to ensure that incidents and accidents do not occur

Figure 4. Typical preliminary risk assessment



within the lifecycle of the product. By implementing a product management safety perspective on system design, the user is provided with the necessary mitigation techniques in order to maintain safety in maintenance and flight operations.

Throughout this chapter, the framework for understanding the need for Human Factors design integration perspectives in the early stages of system design have been provided. There has been the discussion of the need to provide safety management perspectives in the early stages of design in order to reduce the potential for hazards in the maintenance and flight operations environment.

Theoretical perspectives in human factors and safety are important facets in understanding the core of user-system compatibility. However, theoretical perspectives are not the only methods of communicating issues in the industry relative to human factors and safety. Illustrating practical solutions provide the industry with a clearer understanding of the issues pertaining to product system safety and system usability. Three case scenarios related to human factors and system safety engineering are provided below. These case studies will provide the reader with realistic approaches to integrating the human with the system design effectively while providing safety considerations needed to optimize human performance. The case studies are organized as follows:

- Issue
- Effect
- Engineering Analysis
- Recommendations

CASE STUDIES

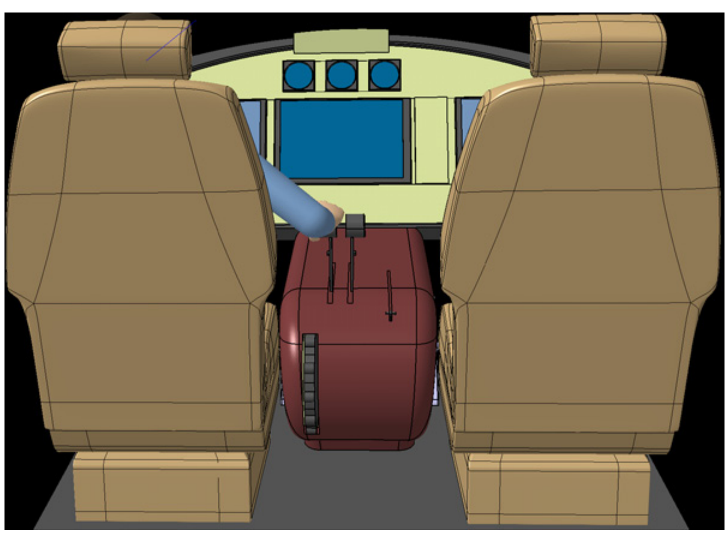
Figure 5.

Case Study# 1

Issues

- The flight crew's inability to reach and read display due to seat configuration (95th percentile male is 1.17 inches away from display and the 5th percentile female is 4.68 inches away).
- Crew's inability to respond to warnings and cautions due to the seat configuration. Program management is concerned with

Figure 5. Crew Station Prototype, National Institute of Aviation Research (NIAR) Laboratories



the ability to meet the aircraft transport category airworthiness standards in 14 CFR 25.1322 (flight crew alerting-warnings, cautions, and advisory lights).

- Need to ensure that the design eye angle position is compatible with user display interface in order to effectively respond to the warning and caution lights. This is also needed to meet the requirements set forth by 14 CFR 25.1321 (arrangement and visibility)
- The display characters and color scheme needs to be evaluated to ensure user system interface compatibility.
    - Need to evaluate the flashing rate of the warnings and cautions on the display

Effects

- May affect the crew's response in the critical phases of flight
- Mission critical and safety critical
- Could affect flight deck certification requirements
- Increased workload

Engineering Analysis

- Determine applicable methods to integrate the human with the system and to meet safety/flight deck certification requirements (Figures 6 and 7).

Case Study #1 Recommendations

- A physical hardware engineering analysis is needed to ensure user system interface compatibility. Change seat position or crew station to accommodate the 5th and 95th percentile population. The seat position change should provide the ability for both percentiles to reach and read the display. If this analysis is not sufficient, HFE many need to provide an analysis based on operator population boundary cases.
- Change seat geometry to accommodate the 5th and 95th percentile population posture. The seat change is needed to ensure the crew members can respond to warnings and cautions annunciations. This change will reduce the system safety risk and increase

Figure 6. Case Study 1: Sevillian, Dujuan (Ph.D. candidate-Cranfield University, UK) Crew Station Human Modeling [Vision Evaluation]-NIAR Laboratories

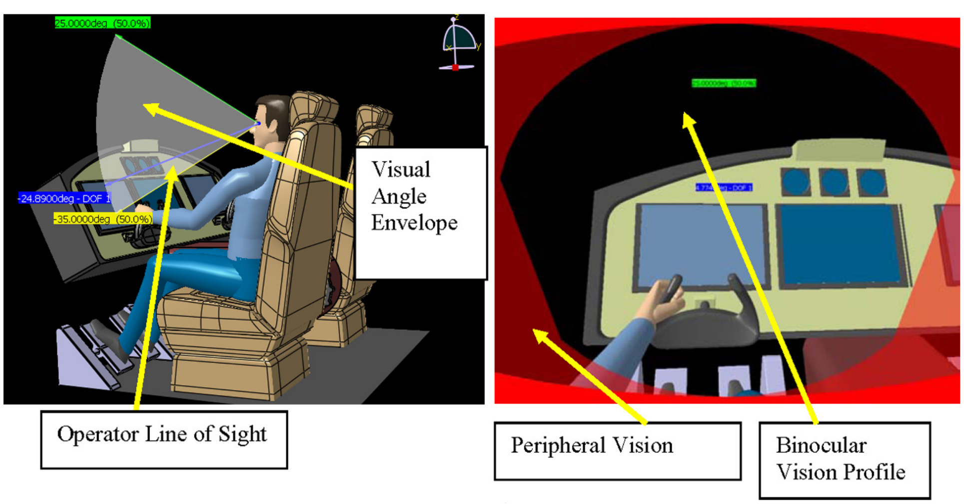
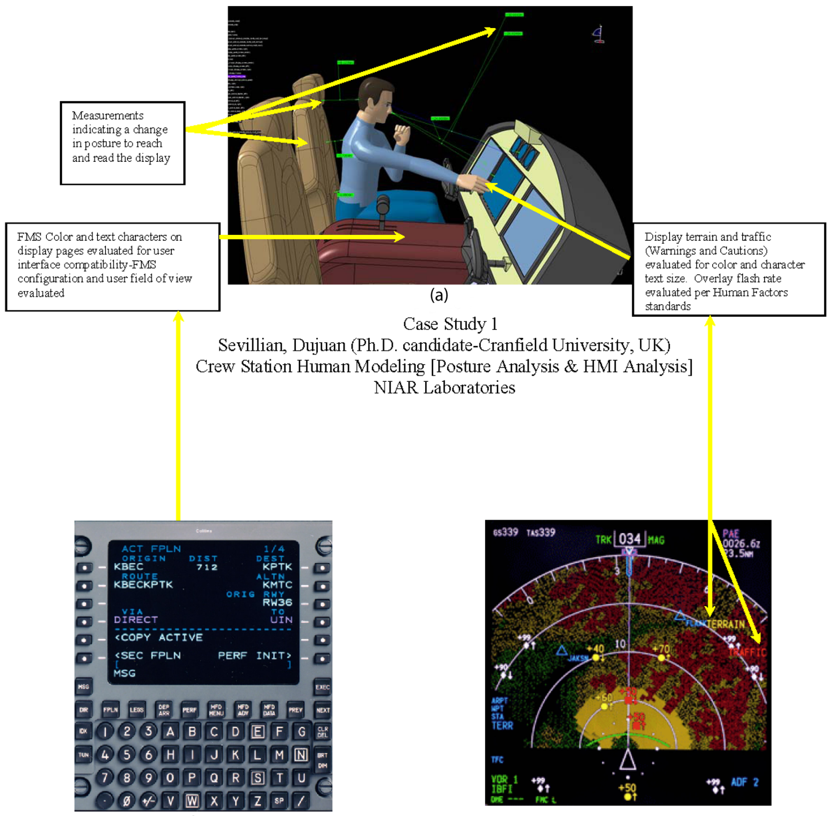


Figure 7.



(a)
Case Study 1
Sevillian, Dujuan (Ph.D. candidate-Cranfield University, UK)
Crew Station Human Modeling [Posture Analysis & HMI Analysis]
NIAR Laboratories

(b)
Flight Management System 3000
© 2010 Rockwell Collins- Used With Permission

(c)
Alert Collision Avoidance System/ Ground Proximity Warning System Display © 2010-Courtesy of Honeywell International Inc.

the likelihood of meeting flight deck certification and airworthiness requirements.

- A cognitive engineering analysis is needed for the Flight Management System (FMS). The display needs to be re-designed to ensure crew members can read the information and interpret the data effectively to ensure mission success. If a software change is not immediately available during flight test operations, a temporary bulletin will need to be implemented in to the FCOM. This is needed to assist the crew members with an alternate means of understanding how to set up the flight plan via the FMS during flight test.
- Need to relocate the FMS system from the throttle quadrant area to an optimal viewing angle to accommodate the human vertical and horizontal visual fields. This analysis is needed to ensure compatibility with user system view angle and to ensure the user can see the FMS system at the optimal viewing distance.
- HFE needs to provide software engineering with standards for display character size and color logic. Also, provide software engineering with display overlay flash rate standards. Last, HFE needs to evaluate the feasibility of integrating aural warnings with the textual terrain and traffic visual

annunciations to provide the user with an increased situational awareness.

Case Study #2

Issues

- Need to ensure the design eye position is evaluated with new seat design, also need to ensure the operator can still see the display
- Need to ensure the applied weight/cushion compression on the seat does not affect the eye position or the seat reference point

Effects

- Inability to meet safety/certification requirements. Crew members may not be able to respond to an emergency effectively
- Safety and mission critical issue
- Could affect the program schedule

Engineering Analysis

- Determine most effective solution to comply with certification and safety requirements (Figures 8 and 9).

Case Study #2 Recommendations

- Ensure seat length and width corresponds with the human physical dimensions. This is needed to ensure compliance with MIL-STD 850.
- Ensure that the change in seat geometry does not affect the eye position or the ability to read and interpret information on the aircrew display.
- Ensure seat geometry change does not affect the ability to meet certification and airworthiness standards.
- Perform a simulation study to determine human body movement limitations within the seat configuration.

Figure 8. Case Study 2: Seat Schematic (MIL-STD 850-Aircrew Station Geometry for Military Aircraft)

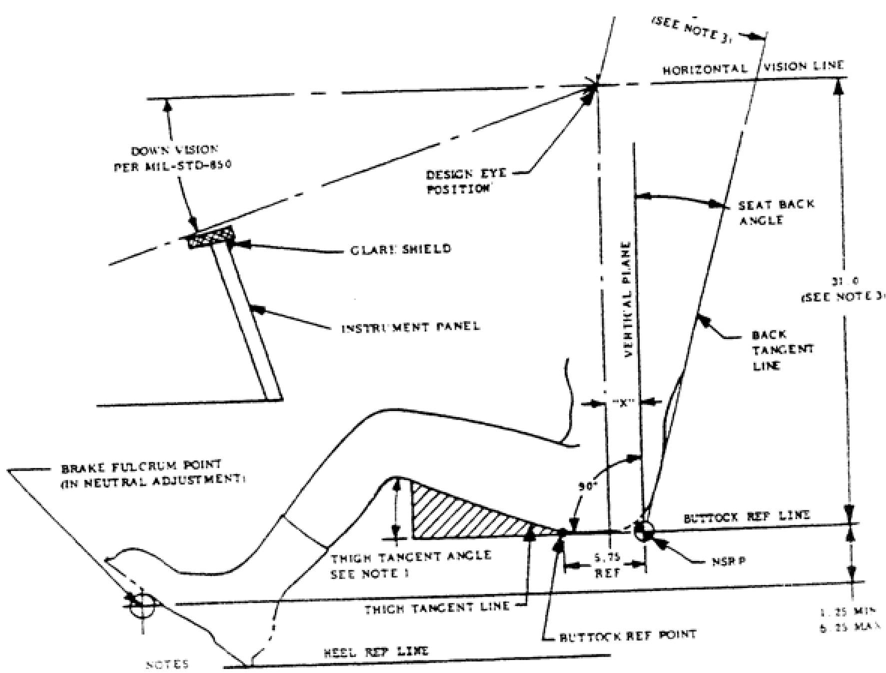
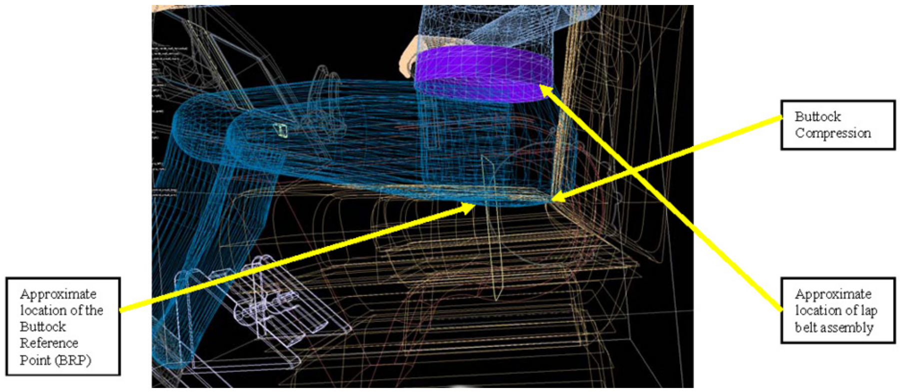


Figure 9. Case Study 2: Sevillian, Dujuan (Ph.D. candidate-Cranfield University, UK) Crew Station Human Modeling [Seat Analysis-Wireframe Configuration] NIAR Laboratories



Case Study #3

Issues

- Flight crew posture and stature inadequate for flight test operations
- Human must remain strapped in lap belt in order to effectively reach rudder pedals but cannot reach pedals - (5^{th} and 95^{th} percentiles issue)

Effects

- Inability to reach rudder pedals
- Possible knee contact/strike with crew station
- Safety and mission critical issue

Engineering Analysis

- Determine most effective way to accommodate crew and meet MIL-STD 1472 and the FAA Human Factors Design Standards (HFDS). Need to utilize MIL-HDBK 759C for anthropometric dimensions (Figures 10 and 11).

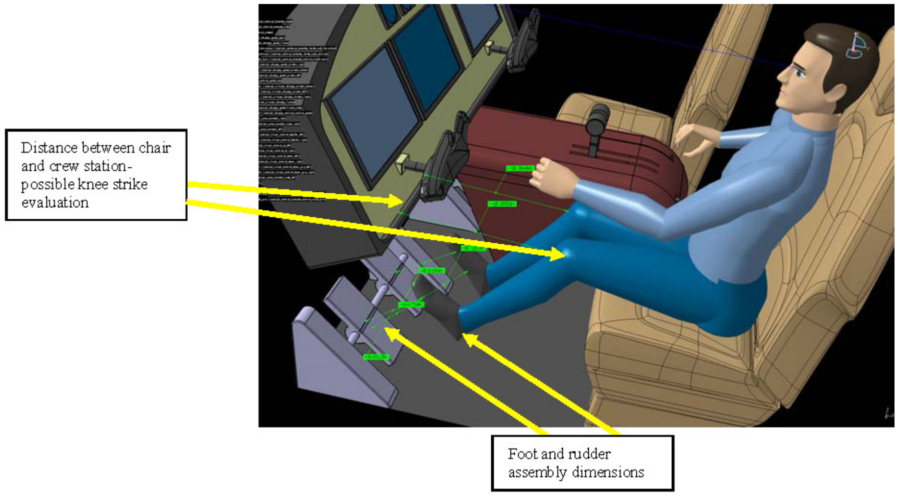
Case Study #3 Recommendations

- Ensure that the dimensions of the human leg and foot integrate with the design of the rudder pedal assembly.
- Ensure the posture of the human is not affected due to the change in rudder pedal configuration.
- Ensure the ability to move seat lateral and vertical to reach rudder pedals. This is needed to accommodate and optimize human abilities to operate the systems within the flight deck.
- Ensure the design configuration does not induce the possibilities of crew member's knees striking the flight deck control panel due to the new engineering design.
- Reduce the system safety risk to an acceptable level for the program and customer.

SOLUTIONS AND RECOMMENDATIONS

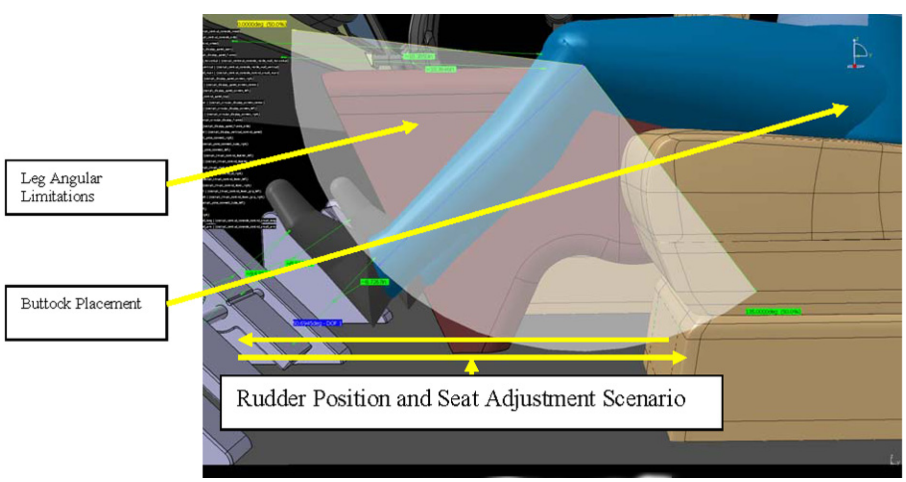
This chapter discussed several methods on human integration philosophies and strategies related to system design. Ensuring effective product safety management is an integral aspect in system design. Risk management is a systematic and collabora-

Figure 10. Case Study 3: Sevillian, Dujuan (Ph.D. candidate- Cranfield University, UK) Crew Station Human Modeling [Rudder/Knee Strike Analysis, Vision, Seat Analysis] NIAR Laboratories



tive effort between program management and the customer which can lead to safe maintenance and flight operations. Implementing effective design solutions provides a practical approach to integrating the human with the system. Understanding program requirements is essential to program success. The program must understand the effects of the requirements on system design. These requirements affect the system design and ultimately the user in their prospective environments. Processes and procedures help facilitate robust strategies for system design management. However, this is not an easy task and it requires a complete understanding of the systems engineering process and how the process interfaces with HFE and safety. Whether the program goal is to implement a HSI program or a HFE program, the program scope and cost dynamics must be further

Figure 11. Case Study 3: Sevillian, Dujuan (Ph.D. candidate-Cranfield University, UK) Crew Station Human Modeling [Rudder and foot dimensions analysis/knee strike analysis, seat analysis] NIAR Laboratories



researched to determine the most effective way to integrate the human with the system. In view of the fact that the overall goal of HFE is to reduce the possibilities of human error in design, strategies must focus on influencing system design concepts with a desire to optimize human system interfaces at the beginning of the program acquisition. In theory, there are seven general recommendations for effective human integration in system design that are recommended. The seven recommendations are as follows:

- Understand the design concept
- Understand the requirements
- Understand the system functions
- Identify the user system interfaces
- Analyze the user system interfaces
- Provide effective human engineering design solutions
- Understand the impact of system design on the user operational environment

FUTURE RESEARCH DIRECTIONS

There is a perpetual need to discover new and more effective methods for integrating the human with the machine and the environment. Research in the field of cognitive engineering appears to be a primary focus within the field of Human Factors Research and Development (HF R&D). New aircraft (unmanned and manned) are continuously being developed and the need to ensure user system compatibility is still a prominent focus.

Pivotal changes in software technology development, such as automation continue to affect human performance in aviation. There have been an abundance of airline and military incidents/accidents related to automation in the flight deck. On June 30, 1994 an A330-300 crashed in Toulouse, France killing seven people. One of the causal factors was the issue related to the user interface with the automation within the flight deck. Automation technology has been a factor influencing the ability for crewmembers to effectively communicate issues within the flight deck. There is a constant need to understand how new automation technology affects Situational Awareness (SA). In particular, CRM in flight deck operations continues to be an issue as more technology is developed for the user community. Questions are asked often by operators, engineers, and scientists regarding automation technology, and the impact on the user interface. 'Does more automation distract the operator'? 'Does more automation provide the operator with the ability to monitor the system better'? Dr. Micah Endsley provides theoretical approaches that can be utilized to understand the impact of automation on the operator interface. According to Endsley, (2000), there are three levels of SA. The three levels are Perception (Level 1 SA), Comprehension (Level 2 SA), and Projection of the Future Situation (Level 3 SA). Understanding these levels provide the framework for determining how to design certain systems that are compatible for the user while increasing SA. There is a constant need to research the impact of new software automation technology on the operator interface during the development and design phases of aircraft systems.

Another focus in HF R&D has been Virtual Reality (VR) and Augmented Reality (AR) applications. Human Factors research in VR & AR environments has provided techniques and strategies to increase human performance. For example, the ability to convey HCI influences on human performance in virtual environments provides the HFE with an understanding of the factors that should be addressed prior to product concept. According to Stanney (1995), an understanding of human sensory and motor physiology is the foundation for understanding the compatibilities and incompatibilities of the human in virtual environments. Human Factors issues such as: Haptic, visual, and auditory perception impact system design in virtual environments (Stanney, 1995). These factors are continuously investigated and are used to determine the impact on human performance.

The book "Technology Engineering and Management in Aviation: Advancements and Discoveries" provides the blueprint for understanding 'how' aviation and aerospace challenges should be managed and executed. Essentially, this chapter illustrates the constant need to effectively manage the human system interface while providing indispensible solutions to development and design.

CONCLUSION

It is common knowledge that air transportation safety is a legitimate reason to design aircraft systems that do not impinge on the user's abilities to perform effectively in particular environments. Designing aircraft systems with the 'Human Element' in perspective provides the foundation for safe maintenance and flight operations. Whether performing maintenance on a landing gear assembly or flying from New York to Japan, it is important that mechanics and pilots have a suitable user interface that enables them to have the ability to effectively communicate and perform their job duties with the highest degree of safety. Therefore, aircraft systems must be developed to optimize the user interface. Expectations of incidents and accidents are likely to occur throughout the lifecycle of a product. However, system designs that optimize human performance can lead to a reduction of human error. Early detection of human performance errors in aircraft design leads to safe and effective maintenance and flight operations.

REFERENCES

Abbott, D. W., & Wise, M. A. (1999). Underpinnings of system evaluation. In Kantowitz, B. H., Daniel, J. A., & Garland, D. J. (Eds.), *Handbook of aviation human factors* (pp. 51–52). Mahwah, NJ & London, UK: Lawrence Erlbaum Associates, Publishers.

Blakey, M. C. (2007, February 15). *Testimony – Statement of Marion C. Blakey.* Retrieved July 19, 2010, from http://www.faa.gov/ news/ testimony/ news_story.cfm? newsId=8184

Chapanis, A. (1996). *Human factors in systems engineering*. New York, NY: John Wiley & Sons.

Defense Acquisition University. (2010, May 17). *Community browser.* Retrieved July 19, 2010, from www.acc.dau.mil/ communitybrowser.aspx? id=314774& lang=en-US

Endsley, M. (1999). Situation awareness in aviation systems. In Kantowitz, B. H., Daniel, J. A., & Garland, D. J. (Eds.), *Handbook of aviation human factors* (pp. 257–259). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Gersh, J. R. (2005). *Cognitive engineering: Understanding human interaction with complex systems*. Baltimore, MD: John Hopkins APL.

Matthews, R. (2004, November). Ramp accidents and incidents constitute a significant safety issue. *ICAO Journal*, 4–25.

McSweeney, J., & Pray, B. C. (2009). *Integration of human factors engineering in to design–An applied approach*. Houston, TX: ABS.

National Highway Traffic Safety Administration. (2008). *Home*. Retrieved July 19, 2010, from http://www-fars.nhtsa.dot.gov/ Main/ index.aspx

National Transportation Safety Board. (1997). *Bus collision with pedestrians*. Washington, DC: National Transportation Safety Board.

National Transportation Safety Board. (2003). *Derailment of Northeast Illinois regional commuter railroad train*. Washington, DC: National Transportation Safety Board.

National Transportation Safety Board. (2007). *Attempted takeoff from wrong runway: Comair Flight 5191*. Washington, DC: National Transportation Safety Board.

National Transportation Safety Board. (2010). *Loss of control on approach, Colgan Air, Inc., operating as Continental connection flight 3407*. Washington, DC: National Transportation Safety Board.

National Transportation Safety Board. (2010, April 8th). *Press release*. Retrieved 19 July, 2010, from http://www.ntsb.gov/ pressrel/ 2010/ 100408.html

Reinach, S. (2007). *An introduction to human systems integration (HSI) in the railroad industry*. Washington, DC: Federal Railroad Adminstration.

Sheridan, T. B. (1993). My anxieties about virtual environments. *Presence (Cambridge, Mass.)*, *2*(2), 141–142.

Shneiderman, B. (1992). *Designing the user interface* (2nd ed.). Reading, MA: Addison-Wesley.

Stanney, K. (1995). *Realizing the full potential of virtual reality: Human factors issues that could stand in the way* (pp. 28–33). Orlando, FL: IEEE.

Strauch, B. (2002). *Investigating human error: Incidents, accidents, and complex systems*. Brookfield, VT: Ashgate Publishing.

Wickens, C. D., Lu, L., & Gordon, S. (1998). *An introduction to human factors engineering*. Prentice Hall.

Wiegmann, D. S. (2004). *HFACS analysis of military and civilian aviation accidents: A North American comparison*. Oklahoma City, OK: ISASI.

Williams, K. W. (2006). *Human factors implications of unmanned aircraft accidents: Flight control problems*. Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

ADDITIONAL READING

Air Traffic control Communications Simulation and Aircrew Training. *Proceedings of the Royal Aeronautical Society, November 5-6, 2003*. London: The Royal Aeronautical Society.

Alexander, A. L., & Wickens, C. D. (2005). *3D navigation and integrated hazard display in advanced avionics: performance, situational awareness, and workload (AHFD-05-10/NASA-05-2)*. Savoy, IL: University of Illinois, Aviation Human Factors Division.

Caldwell, J. A., & Caldwell, J. L. (2003). *Fatigue in aviation*. London: Ashgate Publishing.

Chapanis, A., & Budurka, W. J. (1990). Specifying human-computer interface requirements. *Behaviour & Information Technology*, *9*, 479–492. doi:10.1080/01449299008924261

(1999, June). Does CRM Training Improve Teamwork Sills in the Cockpit?: Two Evaluation Studies. *Human Factors*, *41*(2), 326–343. doi:10.1518/001872099779591169

Endsley, M. R. (2000). Theoretical underpinnings of situational awareness: A critical review. In Endsley, M. R., & Garland, D. J. (Eds.), *Situational Awareness: Analysis and measurement* (pp. 3–32). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Flight Simulation. (2005). *Virtual Environments in Aviation*. Ashgate Publishing, Ltd.

Grady, J. O. (1993). *Systems Requirements Analysis*. New York: McGraw-Hill.

Grose, V. (1987). *Managing risk: Systematic loss prevention for executives*. Upper Saddle River, NJ: Prentice Hall.

Harris, D, & Muir, H. (n.d). *Human factors and aerospace safety: An international journal* [multiple volume and dates]. London: Ashgate Publishing.

Hawkins, F. (1992). *Human factors in flight*. Brookfield, VT: Ashgate Publishing.

Human Error in ATC System Operations. (1980)... *Human Factors*, 535–545.

Jensen, R. (1989). *Aviation psychology*. Brookfield, VT: Gower Publishing.

Johnston, N., Fuller, R., & McDonald, N. (1995). *Aviation psychology: Training and selection*. Brookfield, VT: Ashgate Publishing.

Johnston, N., McDonald, N., & Fuller, R. (Eds.). (1994). *Aviation psychology in practice*. Brookfield, VT: Ashgate Publishing.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*, 1243–1270. doi:10.1080/00140139208967392

Maurino, D., Reason, J., Johnston, N., & Lee, R. (1995). *Beyond aviation human factor*. Brookfield, VT: Ashgate Publishing.

McDonald, N., Johnston, N., & Fuller, R. (1995). *Applications of psychology to the aviation system*. Brookfield, VT: Ashgate Publishing.

National Research Council Committee on Vision. (1983). *Video Displays, Work and Vision*. Washington, DC: National Academy Press.

Norman, S. D., & Orlady, H. W. (Eds.). (1989). *Flight deck automation: Promises and realities. (NASA conference Publication 10036)*. Moffett Field, CA: NASA Ames Research Center.

Nunes, A. & Wickens. C.D. (2005) Clutter effects in CDTI. Unpublished manuscript Podczerwinski, E., Wickens, C.D., & Alexander, A.L. (2002). Exploring the "out of sight, out of mind" phenomenon in dynamic settings across electronic map displays (ARL-01-8/NASA-01-4). Savoy, IL: University of Illinois Aviation Res. Lab.

O'Hare, D., & Roscoe, S. (1990). *Flightdeck performance: The human factor*. Ames: Iowa State University Press.

Proctor, R., & VanZandt, T. (1994). *Human factors in simple and complex systems*. Needham Heights, MA: Allyn and Bacon.

Reason, J. T. (1997). *Managing the risks of organizational accidents*. London: Ashgate Publishing.

Roscoe, S. (1980). *Aviation psychology*. Ames: Iowa State University Press.

Salas, E., Bowers, C., & Edens, E. (2001). *Improving teamwork in organizations*. Lawrence Mahwah, NJ: Earlbaum Associates.

Salas, E., & Maurino, D. (Eds.). (2010). *Human factors in aviation*. New York, NY: Academic Press.

Salvendy, G. (1987). *Handbook of human factors*. New York, NY: Wiley and Sons.

Sanders, M., & McCormick, E. (1993). *Human factors in engineering and design*. New York, NY: McGraw-Hill.

Shappell, S. A., & Wiegmann, D. A. (1997). A human error approach to accident investigation: The taxonomy of unsafe operations. *The International Journal of Aviation Psychology*, *7*, 269–292. doi:10.1207/s15327108ijap0704_2

Simonds, R. H., & Grimaldi, J. V. (1989). *Safety management: Accident cost and control* (4th ed.). New York: McGraw-Hill Professional.

Smith, S. L. (1986). Standards versus guidelines for designing user interface software. *Behaviour & Information Technology*, *5*, 47–61. doi:10.1080/01449298608914498

Strauch, B. (2002). *Investigating human error: Incidents, accidents, and complex systems*. Brookfield, VT: Ashgate Publishing.

Thygerson, A. L. (1992). *Safety (Rev)*. Boston: Jones & Bartlett Publishers.

Wagenaar, W. A., & Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies*, *27*(5-6), 587–598. doi:10.1016/S0020-7373(87)80017-2

Wickens, C., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: Harper Collins.

Wickens, C. D., Gempler, K., & Morphew, E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors.*, *2*(2), 99–126. doi:10.1207/STHF0202_01

Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). Implications of graphics enhancements for the visualization of scientific data: dimensional integrality, stereopsis motion, and mesh. *Human Factors*, *36*(1), 44–61.

Wiener, E., & Nagel, D. (Eds.). (1988). *Human factors in aviation*. San Diego, CA: Academic Press.

Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. (NASA contractor Report 177528)*. Moffett Field, CA: NASA Ames Research Center.

Woodson, W. E., Tillman, B., & Tillman, P. (1992). *Human Factors Design Handbook* (2nd ed.). New York: McGraw-Hill.

Yacavone, D. (1993). Mishap trends and cause factors in Naval aviation: A review of Naval Safety center data, 1986-1990. *Aviation, Space, and Environmental Medicine*, *64*, 392–395.

KEY TERMS AND DEFINITIONS

Critical Design Review: (CDR): The CDR is the final review before the production stage of the system. The purpose of this review is to compile all of the details regarding software and hardware configurations.

Federal Aviation Administration: (FAA): U.S. Government agency charged with developing regulatory standards providing regulatory oversight to the aviation aerospace industry.

Human Factors Engineering (HFE): Is the application of human capabilities and limitations as it relates to the design of systems.

Human Factors: Human Factors are the factors influencing the behavior of the human based on several dynamic critical interfaces.

Human Systems Integration (HIS): Is the systematic process of integrating the human with the system so that the human has the ability to operate, maintain, and support personnel needs within the applicable environment.

Information Processing Model: (IPM): The IPM is a model that portrays how the human interprets a stimulus and responds to certain system inputs in a particular environment.

Line Replaceable Unit (LRU): A system designed to be removed and replaced expeditiously.

National Transportation Safety Board: (NTSB): Agency charged by U.S. Congress to investigate all U.S. transportation related accidents.

ORR: The ORR is a review that determines the feasibility in proceeding to the concept exploration.

Preliminary Design Review: (PDR): The PDR is a review with the customer, program management, and various engineers to review the current design philosophy. The PDR includes the approach to various design configurations such as hardware component specification, and detailed requirements for future test evaluations.

System Design Review: (SDR): The SDR is a review developed to determine if the integrated system will meet the requirements.

System Requirements Review (SRR): The SRR is conducted with the customer to understand how the requirements will be utilized to design the system.

Test Readiness Review: (TRR): The TRR is conducted to review the design integration as it relates to the ability for the system to operate effectively in the operational environment.

This work was previously published in Technology Engineering and Management in Aviation: Advancements and Discoveries, edited by Evon Abu-Taieh, Asim El Sheikh and Mostafa Jafari, pp. 73-100, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 49

Adoption of Information Technology Governance in the Electronics Manufacturing Sector in Malaysia

Wil Ly Teo
Universiti Teknologi Malaysia

Khong Sin Tan
Multimedia University, Malaysia

ABSTRACT

Past studies and surveys of top management in business and information technology (IT) have shown the importance of strong IT governance in delivering results to the business. This research investigates the extent to which empirical results from past studies is applicable to the electronics manufacturing sector in Malaysia. Empirical evidence from 33 organisations in this sector indicates that having the right decision owners making appropriate decision types lead to better IT governance performance. Organisations with growth as their primary goal demonstrate marginally higher IT governance performance, contrary to expected outcomes. The research also shows that awareness of industry IT governance frameworks is not related to IT governance performance. We conclude that adoption of IT governance is on track, though familiarity with technicalities of the available frameworks should be improved.

INTRODUCTION

Overview

Governance, in general terms, can be understood as the act or process of governing, specifically on the authoritative direction or control. The term "IT governance" as it is known today surfaced in the early 1990s with Loh and Venkatraman (1992) and Henderson and Venkatraman (1993) using the term to describe the mechanisms to ensure attainment of necessary IT capabilities.

In this paper, the terms IS and IT are used interchangeably. The definition of IT governance in this context is based upon the following:

DOI: 10.4018/978-1-4666-1945-6.ch049

1. Van Grembergen (2002) terms IT governance as the role of Board, executive management and IT management to achieve alignment of business and IT strategies through control over formulation and implementation of IT strategy.
2. IT Governance Institute (2003) defines IT governance as "the responsibility of the board of directors and executive management" and "is an integral part of enterprise governance and consists of the leadership and organisational structures and processes that ensure that the organisation's IT sustains and extends the organisation's strategies and objectives" (p. 10).
3. IT governance is "specifying the decision rights and accountability framework to encourage desirable behaviour in the use of IT" (Weill & Ross, 2004, p. 8).

Research Context

Various studies and surveys of top management from business and IT have shown the importance of strong IT governance in delivering results to the business. Of significant importance is the study of IT governance in 256 companies in 23 countries conducted by Weill and Ross (2004). The study provided empirical evidence that strong IT governance leads to more successful business outcomes. This research intends to validate the empirical evidence in the context of electronics manufacturing sector in Malaysia. The motivating factors behind such a research context will be discussed shortly.

Malaysian Industry Development Authority reported that the electronics industry (in combination with electrical industry) achieved a 10.8% increase in sales value to RM195.7 billion. The production index of the industry expanded by 8.7%. Specifically for the electronics industry, the production of semiconductors and other electronic components expanded by 10.4% in tandem with the growing global semiconductor market (Malaysian Industrial Development Authority, 2006).

In addition to the economic significance of this sector to the national economy, there are two other factors behind the choice of research context. Firstly, distinguishing from financial sector where IT governance implementation is often mandated by statutory requirements, such initiatives in the electronics manufacturing sector are likely to be voluntary and driven by business benefits.

Secondly, the use of technology, including IT, is pervasive as the industry moves up the value chain after more than three decades since the first foreign investment started operation locally. The maturity of IT usage means the focus has shifted from basic infrastructure initiatives towards effective use of IT, hence more likely to have emphasis on elements of IT governance.

Problem Statement

With the motivating factors described above, this research was carried out in late 2006 to validate the empirical results from Weill and Ross (2004) in the organisational context of electronics manufacturing sector in Malaysia. Specifically, the research questions are:

1. What is the as-is IT governance level?
2. Why do organisations implement IT governance?
3. How successful has IT governance contributed to the intended outcomes?
4. What are the factors affecting IT governance performance?

The research intends to explore the as-is situation of structure, awareness and performance of IT governance, understand the motivation of IT governance initiatives and assess the alignment of IT governance with business strategies.

Significance of the Study

This research is beneficial to academic researchers, practitioners and policy makers who are interested to understand the as-is situation of IT governance level in electronics manufacturing industry in Malaysia. There is limited prior academic research on IT governance in Malaysia, particularly empirical-type and within this industry. This paper extends the earlier preliminary conference paper (Tan, Eze, & Teo, 2008) in the sense of depth and coverage of the results.

Organisation of the Paper

The relevant theoretical background in IT governance especially in aspects related to this research will be briefly covered, followed by research design, data analysis and discussion.

LITERATURE REVIEW

Boundaries of IT Governance

IT Governance Institute (2003) describes the two fundamental concerns of IT governance as:

- IT's delivery of value to the business, driven by strategic alignment of IT with the business
- Mitigation of IT risks, embedding accountability into the enterprise

In a general sense, Sohal and Fitzpatrick (2002) distinguish governance as the creation of a setting in which others can manage effectively, from management domain where operating decisions are made. Taking this distinction to the IT field, hence IT management focuses on operational aspects, whereas IT governance deals with strategic issues (Van Grembergen, 2003).

Each day, IT managers are faced with a myriad of issues requiring them to make decisions. To remain focused on issues of strategic importance, it is necessary to distinguish:

- Key strategic issues and daily operational issues.
- IT-relevant issues and the business domain

The key IT decision model from Weill and Ross (2004) classifies IT decisions into 5 areas:

- IT principles
- IT architecture
- IT infrastructure
- Business applications needs
- IT investment and prioritisation

IT Principles

IT principles are the high-level statements about how IT is used in the business (Weill & Ross, 2004) or referred to as maxims by Broadbent and Kitzis (2004). Organisations are often confronted with two contradictory dilemmas (Kayworth & Sambamurthy, 2000). IT is expected to respond quickly and flexibly within individual business units. At the same time, enterprise-wide integration depends on the ability of IT to exploit cross-unit synerties. A set of IT principles will clarify the priorities consistent with the organisation's goals.

IT Architecture

Solution architecture addresses both business processes and technological aspects of solution implementation. Effective information architecture is the answer to the dual challenge of "speed and flexibility" and "low cost and efficiency (Allen and Boynton, 1991). Consistent execution is a benefit of standardisation, which in turn results predictability and efficiency.

IT Infrastructure

IT infrastructure is the foundation of planned IT capabilities available throughout the business as shared and reliable services and used by multiple applications (Weill, Subramani, & Broadbent, 2002). The study by Weill, Subramani and Broadbent (2002) found significant correlation between strategic agility and IT infrastructure capability. Managers use IT principles to identify IT infrastructure services that enable the organisation to achieve competitive advantage.

Business Application Needs

The business value of IT is most directly delivered through fulfilment of specific business needs. However, identification of business needs for IT applications often has two conflicting objectives of being creative while remaining disciplined (Weill & Ross, 2004). Both business and IT need to have a common agreement that new functionality might need to be sacrificed if they cause deviation to architectural principles.

IT Investment and Prioritisation

IT investment and prioritisation process determines the amount to be spent on IT, identifies the IT initiatives to be undertaken, and reconciles conflicting needs of individual business units. There is no specific framework to determine the amount to be spent on IT. Departing from the conventional business case and discounted cash flow techniques on IT investment decisions, Ross and Beath (2002) and Irani and Love (2002) propose alternative approaches to justify IT investments and expenditures.

Identifying the Decision-Makers

The study of IT governance from the perspective of who has the power to make decisions is as old as the use of computer systems itself. The simplest form of such study considers only either centralised or decentralised decision-making structure began with Garrity (1963) and continued into late 1990s in various forms (Boynton & Zmud, 1987; Brown, 1997; Olson & Chervany, 1980).

The expansion into vertical form considers multiple degrees of centralisation/decentralisation in the form of continuous classification (Ein-Dor & Segev, 1978; King, 1983; Olson & Chervany, 1980; Tavakolian, 1989), discrete nominal classification and redefinition of extremes within a decentralised form. (Boynton & Zmud, 1987; Dixon & John, 1989; Rockart, 1988). In addition to that, horizontal expansion considers varying degrees depending on the type of decision, rather than the entire IT organisation (Sambamurthy & Zmud, 1999). In this area of study, the focus was on the use of technology (Allen & Boynton, 1991; Brown & Magill, 1994; Dixon & John, 1989; Zmud, Boynton, & Jacobs, 1986).

The pragmatic approach taken in this research follows the ownership of decision-making classification by Weill and Ross (2004, p. 27):

- Business Monarchy: A group of business executives or individual executives (CxOs). Includes committees of senior business executives (may include CIO). Excludes IT executives acting independently.
- IT Monarchy: Individuals or groups of IT executives
- Feudal: Business unit leaders, key process owners or their delegates
- Federal: C-level executives and business groups (e.g. business units or processes); may also include IT executives as additional participants. Equivalent of the central and state governments working together.
- IT Duopoly: IT executives and one other group (e.g., CxO or business unit or process leaders)
- Anarchy: Each individual user

The Execution Stage

Formal Governance Mechanisms

Horizontal mechanisms encourage contacts between individuals in order to coordinate the work of two units (Brown, 1999). These can be task-specific groups (e.g. ad hoc task forces), permanent teams with cross-unit representatives or formal roles responsible for collaboration across units. Commonly found mechanisms include steering committee with membership from IT or mixture of both business and IT, account management teams, and programme and project office (Weill & Ross, 2004).

Once decisions are made, monitoring and periodic reviews are necessary to ensure the execution is on the right path. The mechanisms for monitoring execution include standard operation management processes, service level agreements (SLA), organisation-wide project management methodology and key performance indicators or balanced scorecard (Grembergen & De Haes, 2005; Van Grembergen & Van Bruggen, 1997).

Communication includes formal announcements, information channels and education efforts that propagate IT governance principles, policies and as reflected in the outcome of IT decision-making proceses (Weill & Ross, 2004). In this research, only formal communication mechanisms are considered whereas soft aspects of personal contacts and relationships between the IT management and business are excluded.

Governance Frameworks

There are various frameworks related to IT governance in one way or another. The three common governance frameworks are briefly described here.

Control Objectives for Information and related Technology (COBIT) was first released in 1996. COBIT is organised into four domains: Planning and Organisation, Acquisition and Implementation, Deployment and Support, and Monitoring, 34 high-level processes and 210 control objectives (IT Governance Institute, 2007).

IT Infrastructure Library (ITIL) was developed in the late 1980s by the UK's Office of Government Commerce (OGC). ITIL is intended to increase IT business effectiveness, reduce costs and improve IT services. IT service management practices from ITIL are formalised as the British Standard 15000 and later followed by ISO 20000:2005 Standard (Office of Government Commerce, 2007).

Capability Maturity Model Integration (CMMI) for software development was established in 1986 at Carnegie Mellon University by the Software Engineering Institute (SEI). CMMI defines five stages of organisational maturity with respect to software development. The objective is to improve predictability, effectiveness and control of an organisation's software processes (Software Engineering Institute, 2005).

Effects of Organisational Context

A framework of organisational context predicting the success and failure of IT was proposed by Ein-Dor and Segev (1978). However, no empirical evidence was provided to test the validity of the framework.

Empirical studies relating organisation context and IT governance finally emerged in 1980s through 1990s. These studies were mostly done on organisations in the United States (Clark, 1992; Ein-Dor & Segev, 1982; Olson & Chervany, 1980; Tavakolian, 1989), Europe (Earl, 1989, as cited in Brown, 1997) and Israel (Ahituv, Neumann, & Zviran, 1989).

The most commonly studied organisational context variables were:

- Organisation size (employees or revenue)
- Governance form
- Competitive and business strategy
- Diversity in business
- Industry

Organisation Size

Olson and Chervany (1980) found no differences by organisation size. Interestingly, Ein-Dor and Segev (1982) discovered inverse relationship between organisation size and degree of centralisation (however this study was limited to one city in the United States).

Governance Form

Ein-Dor and Segev (1982) found positive relationship between organisation structure and IT governance. Such finding was also found to be true for organisations in Israel where the only significant determinant in distributing computing resources was the decision-making process of the organisation (Ahituv et al., 1989).

Competitive and Business Strategy

Tavakolian (1989) studied competitive strategy (defender, prospector, and analyser) as the organisational variable relating to IT governance. The study found that companies with defender strategy have more decentralised structure of IT governance compared with prospector strategy. This was extended by Henderson and Venkatraman (1993) on the range of competitive strategy to what was termed as "fundamental domains of strategic choice". The four strategic choices identified in the study were business strategy, information technology strategy, organisational infrastructure and processes, IT infrastructure and processes.

Diversity in Business

Business synergy is found not to be a dominant contingency for IT governance mode or IT synergies of multi-business firms (Sambamurthy & Zmud, 1999). However, Tanriverdi (2006) argues that IT resources can create an IT-based coordination mechanism in a multi-business firm and enhance boundary-spanning organisational capabilities that exploit cross-unit business synergies.

Industry

Of all the organisational context variables discussed, industry is the only one with common agreement by authors of past studies. Olson and Chervany (1980), Ahituv, Neumann and Zviran (1989) and Clark (1992) found no relationship between industry and IT governance. Furthermore, Brown and Magill (1994), in their study of predictors for movement from one to another form of governance, found that industry type is not a strong predictor.

RESEARCH METHODOLOGY

Construct

The theoretical framework is built upon the three key considerations of IT governance outlined by (Weill & Ross, 2004) discussed in the theoretical section: what decisions should be made, who should make these decisions, and how will these decisions be made and monitored.

The questions of "what" and "who" lead to ownership of decision-making as the first construct, while "how" naturally leads to formal governance mechanisms as another construct.

Construct 1: Ownership of decision-making
Construct 2: Formal governance mechanisms

Further, it is imperative that we do not lose sight of the fundamental concerns of IT governance: strategic alignment of IT with the business, and mitigation of IT risks. Thus, the discussion of IT governance is incomplete without considering the business context in which IT operates.

Construct 3: Organisational context

Finally, these 3 constructs lead to the outcome of IT governance.

Construct 4: IT governance performance

Variables

IT Decisions

Each of the decisions can be made by any of the following groups of decision makers, as classified by Weill and Ross (2004).

Decision Types:

- IT principles
- IT architecture
- IT infrastructure
- Business application needs
- IT investment and prioritisation

Decision Makers:

- Business Monarchy
- IT Monarchy
- Feudal
- Federal
- IT Duopoly
- Anarchy (Weill & Ross, 2004, p. 59)

The resultant variable set consists of five variables (one for each type of decision). The value of each variable is reclassified from the six categories listed above into three broad categories, i.e. business, IT or joint decisions. Reason for this reclassification will become apparent in the section on formulation of hypotheses.

Taking into account that one group might not be consistently used for the same type of decision, a 5-point Likert scale representing frequencies from "Never" to "Always" is used.

Variable 1a: Extent of joint business-IT involvement in decision-making on IT principles (scale)

Variable 1b: Extent of independence of IT in decision-making on IT architecture (scale)

Variable 1c: Extent of independence of IT in decision-making on IT infrastructure (scale)

Variable 1d: Extent of joint business-IT involvement in decision-making on business application needs (scale)

Variable 1e: Extent of independence of business in decision-making on IT investment and prioritisation (scale)

Formal Governance Mechanisms

The variables used to measure this construct are derived from the three aspects of making and monitoring IT decisions, again basing upon (Weill & Ross, 2004, p. 86):

- Coordinating the decision-making process.
- Monitoring to ensure decisions are executed as agreed upon.
- Communicating constantly with stakeholders.

Similar to the previous set of variables, a 5-point Likert scale representing frequencies from "Never" to "Always" is used to take into account the fact that a mix of different mechanisms is possible for each aspect.

Variable 2a: Extent of usage of formal governance coordination of decision-making mechanisms (scale)

Variable 2b: Extent of usage of formal governance execution mechanisms (scale)

Variable 2c: Extent of usage of formal governance communication mechanisms (scale)

In addition, one of the research objectives is to "explore the as-is situation of structure, *awareness* and performance of IT governance". The extent of awareness is measured as awareness of industry of frameworks related to IT governance, which in turn implies clarity of direction taken by the IT organisation with respect to IT governance.

Variable 2d: Extent of awareness of industry frameworks related to IT governance (scale)

Organisational Context

In this research, the organisational context variables of interest are organisation size as characterised by revenue and strength of workforce, and the primary goal of an organisation (profit or growth orientation)

Variable 3a: Revenue (ordinal)
Variable 3b: Strength of workforce (ordinal)
Variable 3c: Primary goal of organisation (nominal)

Governance Performance

Organisations have different expected outcomes for implementing IT governance. The objectives could be one or more of the following:

- Cost-effective use of IT
- Effective use of IT for growth
- Effective use of IT for asset utilisation
- Effective use of IT for business flexibility (Weill & Ross, 2004, p. 121)

Governance performance is a weighted average of the importance of outcome and contribution of IT governance to achieve them.

Variable 4: IT governance performance (scale) is calculated using the equation in Box 1.

Theoretical Framework

Figure 1 depicts the relationship between variables constituting the theoretical framework.

Hypotheses

Ownership of Decision-making

IT principles are defined as basic philosophies about IT, articulated through a set of IT management principles that summarise how the company would use IT to achieve its goals. Davenport, Hammer and Metsisto (1989) argue that keeping with the corporate strategy means technology decisions must be guided by agreed principles.

However, principles are by no means static statements. Periodic review is required, not only for alignment with changing corporate strategy, but also to harness new possibilities arising from technology development. Therefore, it is imperative that both business and IT are jointly involved in the formulation and review of IT principles, both from perspective of corporate strategy and technology development.

H1a: Joint decision-making by business on IT for IT principles is correlated with high governance performance.

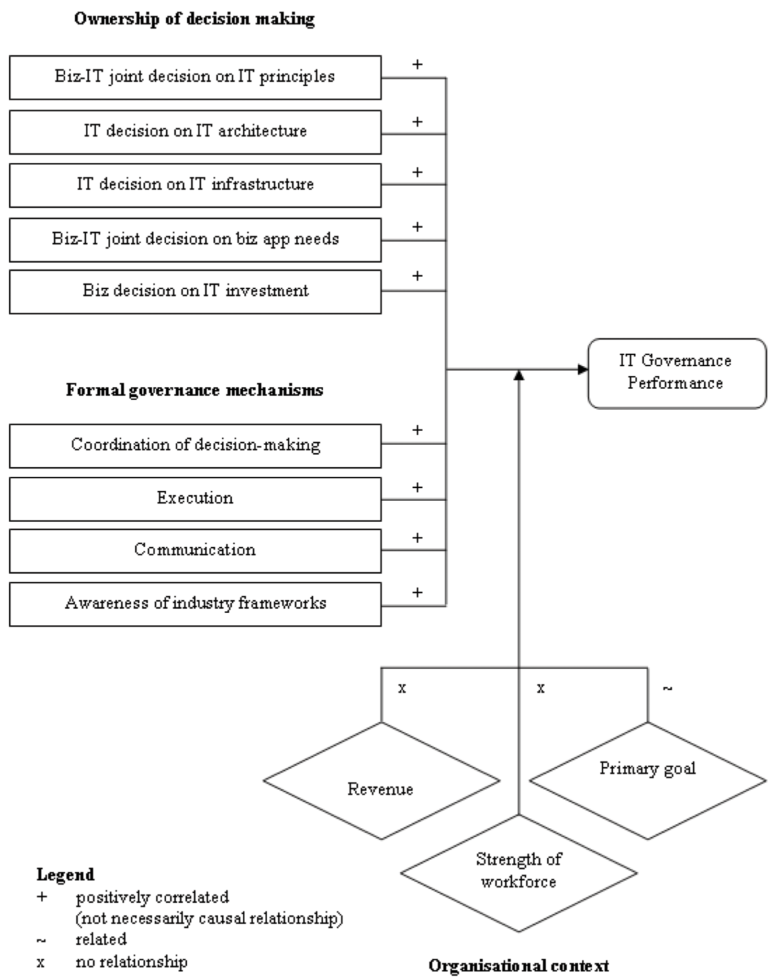
Business process integration and standardisation are key responsibilities of the process

Box 1.

$$\frac{\sum_{n=1 \text{ to } 4}\left(\text{importance of outcome X influence of IT governance}\right)}{\sum_{n=1 \text{ to } 4}\left(\text{importance of outcome}\right)}$$

(Weill & Ross, 2004, p. 240)

Figure 1. Theoretical framework



owner from the business side. A pre-requisite for integration and standardisation is the formulation of IT architectural decisions. These provide the foundation for IT capabilities of the organisation (Weill & Ross, 2004). Similarly, IT infrastructure capability is a key contributor to strategic agility (Weill et al., 2002).

Although IT architecture and infrastructure are based upon requirements articulated as part of IT principles, these are technical decisions within the domain of IT specialists. Therefore, effective decision-making is best left to IT specialists guided by objective criteria, independent of input from non-specialists.

H1b: Sole decision-making by IT on IT architecture is correlated with high governance performance.

H1c: Sole decision-making by IT on IT infrastructure is correlated with high governance performance.

Creativity in delivering customer value through IT and disciplined execution consistent with architectural principles are two key considerations in deciding business application needs. As such, both business and IT have to agree on the trade-off between fulfilment of business needs and compliance with architectural principles.

H1d: Joint decision-making by business and IT on business application needs is correlated with high governance performance.

The IT investment justification model proposed by Ross and Beath (2002) categorises investments as one of transformation, renewal, process improvement or experiments. Except for renewal, the other three categories of IT investment are usually directly driven by business needs, including the need to support a new business model, improve operational performance and enable new technologies or ideas.

Organisations undertake technology renewal to reduce cost or raise quality of IT services. In addition, end of product life or vendor ceasing their services are two externally-driven reasons to embark on technology renewal. The IT organisation's annual budget allocation is usually used to fund renewal activities. In spite of this, we must not overlook the fact that such allocations are decided by business, considering financial standing of the organisation and input from IT with regards to mandatory renewal activities.

H1e: Sole decision-making by business on IT investment is correlated with high governance performance.

Formal Governance Mechanisms

Good IT governance does not happen by chance. Therefore formal governance mechanisms are necessary to achieve it.

H2a: Organisations with formal governance coordination of decision-making mechanisms have high governance performance.
H2b: Organisations with formal governance execution monitoring mechanisms have high governance performance.
H2c: Organisations with formal governance communication mechanisms have high governance performance.

Rising awareness of industry frameworks related to IT governance, such as COBIT, ITIL and CMMI, is expected to lead to widespread activities to improve IT governance. Availability of these industry-wide frameworks means organisations starting out on this journey do not need to reinvent the wheel and reduce implementation risks, thereby increasing chances of success.

H2d: Awareness of industry frameworks related to IT governance is positively correlated with governance performance level.

Organisational Context

IT governance is not limited to large corporations, just as corporate governance applies to all joint-stock corporations. IT governance affects big and small, profit and non-profit organisations. The difference lies in how rigorous the implementation is, but not how successful IT governance contributes to the intended outcome.

H3a: IT governance performance is not affected by annual revenue (organisation size).
H3b: IT governance performance is not affected by strength of workforce (organisation size).

In the quest for a new customer base and product innovation, growth-oriented organisations require extremely agile IT systems and flexible IT services to deal with fast-changing business requirements. On the other hand, organisations with profit as the primary goal strive to maximise profit through operational efficiencies and economies of scale through standardised and tightly controlled processes and tools. This could be translated to a more immediate improvement in IT governance performance.

H3c: Organisations with profit as the primary goal have a high IT governance performance compared to organisations with growth as the primary goal.

Research Design

This research engages in hypotheses testing. It attempts to explain the relationship between ownership of IT decisions, formal governance mechanisms, organisational context and IT governance performance. In this research, data collection is done mainly via survey. In administering the survey, minimal researcher interference is applied, limited to providing clarification to respondents where necessary. Every company in the survey forms a unit of analysis.

Data Collection

This research is carried out using a set of questionnaire structured in 5 sections:

- **Section A:** Organisational context
- **Section B:** IT decision-making
- **Section C:** Coordination, monitoring and communication mechanisms
- **Section D:** IT governance framework
- **Section E:** IT performance governance

Section A uses dichotomous and interval-scale. For sections B and C, 5-point Likert scale is used (1=Never, 2=Rarely, 3=Sometimes, 4=Often and 5=Always). The 4-point Likert scale is used for section D (1=Currently Used, 2=Intend to Use, 3=No Intention to Use, and 4=Never Heard of). Finally Section E uses two different 5-point Likert scales are used for different questions in this section (1=Totally Not Important, 2=Not Important, 3=Quite Important, 4=Important and 5=Very Important; 1=Totally Not Successful, 2=Not Successful, 3=Quite Successful, 4=Successful and 5=Very Successful). Respondents are also provided the opportunity to provide response in free text.

Respondents have the option to complete the survey through one of the following methods:

- Web-based survey
- E-mailing the completed questionnaires
- Posting the completed questionnaire using the postage-paid reply envelope

Unique identification numbers are attached to survey invitations to ensure no duplicates in data collection.

Sampling

The population of this research is determined primarily using the member directory of the Federation of Malaysian Manufacturers (FMM), as well as through industry contacts. FMM represents more than 2,000 manufacturing and industrial service companies of various sizes in Malaysia (Federation of Malaysian Manufacturers, 2006).

The organisations included in this research are involved in the manufacturing of one or more of the following products: diode, electronic, integrated circuit, memory product, resistor, semiconductor, sipmos, transistor, and wafer. Care was taken to clean the population frame by removing obvious mismatches (e.g. “wafer” as a product of the food industry, as opposed to “wafer” fabrication in the semiconductor industry).

The finalised population frame consisted of 84 organisations. No sampling was done as the number is relatively small. All organisations in the list were invited to participate in the survey which ran for approximately one month.

DATA ANALYSIS

Sample Characteristics

Out of 84 surveys distributed, 30 respondents answered the questionnaires online, 3 responded by returning the completed paper-based survey form through postal mail, while 2 declined to participate citing confidentiality as a reason. 33

usable survey responses were collected which represented almost 40% response rate (N = 33).

Reliability Analysis

The questionnaire section on ownership of decision-making consists of five main questions, each corresponding to IT principles, IT architecture, IT infrastructure, business application needs and IT investment and prioritisation. Each main question consists of seven sub-questions (items) representing possible combinations of the decision owner. Respondents were asked to rate how often each group makes the decision. Reliability analysis for inter-item consistency of the recoded answers is tabulated in Figure 2. Cronbach's Alpha value of 0.7 is used as reference value for assessing item consistency (Nunnally, 1978).

Descriptive Statistics

Ownership of Decision-Making and Formal Governance Mechanisms

The descriptive statistics are shown in Figures 3 and 4 respectively.

Organisational Context

The distribution of annual revenue (variable 3a) is skewed towards large corporations with significant representations of small and medium-sized companies. Similarly, the distribution of strength of workforce (variable 3b) is also skewed towards medium-sized companies and large corporations having more than 100 and 1000 employees respectively. Finally, distribution of primary goals of the companies (variable 3c) is almost equally distributed between profit and growth.

Figure 2. Inter-item consistency

Item	Cronbach's Alpha	Remarks
IT principles	0.662	Marginally lower (acceptable)
IT architecture	0.694	Marginally lower (acceptable)
IT infrastructure	0.722	2 cases were found to contribute to initial low value of 0.455, revised calculation resulted in acceptable value of 0.722.
Business application needs	0.722	
IT investment and prioritisation	0.702	

Figure 3. Descriptive statistics for ownership of decision-making

Variable	Scale	Mean	Standard Deviation	Range
1a	1–5	3.36	0.30	3.00 – 3.86
1b		2.99	0.33	2.14 – 3.43
1c		3.09	0.27	2.57 – 3.57
1d		3.40	0.38	2.86 – 3.86
1e		3.19	0.38	2.57 – 4.29

Figure 4. Descriptive statistics for formal governance mechanisms

Variable	Range	Mean	Standard Deviation	Range
2a	1–5	3.47	0.58	1.40 – 5.00
2b		3.62	0.63	2.00 – 5.00
2c		4.12	0.44	3.67 – 5.00
2d	1–4	2.09	0.60	1.00 – 3.00

IT Governance Performance

IT governance performance is measured as how important the outcomes are to them and how successful IT governance has contributed to each outcome. The mean IT governance performance is 3.42 with a standard deviation of 0.44 (on a scale of 1 to 5). The values range from 3.00 to 4.56.

Inferential Statistics

Correlation between Ownership of Decision-Making and Governance Performance

Pearson Correlation test shows significant correlation between the abovementioned variables (p-value of less than 0.01):

- Extent of joint business-IT involvement in decision-making on IT principles
- Extent of independence of IT in decision-making on IT infrastructure
- Extent of joint business-IT involvement in decision-making on business application needs
- Extent of independence of business in decision-making on IT investment and prioritisation

The extent of independence of IT in decision-making on IT architecture showed significant correlation with IT governance performance (p-value of less than 0.05). Details of inferential statistics are shown in Figure 5.

Therefore, the following hypotheses pertaining to ownership of decision-making and governance performance are not rejected.

H1a: Joint decision-making by business on IT for IT principles is correlated with high governance performance.
H1b: Sole decision-making by IT on IT architecture is correlated with high governance performance.
H1c: Sole decision-making by IT on IT infrastructure is correlated with high governance performance.
H1d: Joint decision-making by business and IT on business application needs is correlated with high governance performance.
H1e: Sole decision-making by business on IT investment is correlated with high governance performance.

Correlation Between Formal Governance Mechanisms and Governance Performance

Formal governance mechanisms for communicating IT issues also shows very significant correlation with IT governance performance (p-value of less than 0.01), while formal governance mechanisms for coordinating IT decision-making activities and monitoring execution of IT decisions show significant correlation with IT governance

Figure 5. Pearson Correlation between ownership of decision-making and governance performance

Correlations

		IT principles	IT architecture	IT infrastructure	Business application needs	IT investment and prioritisation	IT governance performance
IT principles	Pearson Correlation	1	-.047	.243	.663**	-.075	.483**
IT architecture	Pearson Correlation	-.047	1	.876**	.105	.282	.348*
IT infrastructure	Pearson Correlation	.243	.876**	1	.210	.238	.505**
Business application needs	Pearson Correlation	.663**	.105	.210	1	-.139	.509**
IT investment and prioritisation	Pearson Correlation	-.075	.282	.238	-.139	1	.494**
IT governance performance	Pearson Correlation	.483**	.348*	.505**	.509**	.494**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 6. Pearson Correlation between formal governance mechanisms and governance performance

Correlations

		Coordinate IT decision-making activities	Monitor execution of IT decisions	Communicate IT issues	IT governance framework	IT governance performance
Coordinate IT decision-making activities	Pearson Correlation	1	.813**	.420*	.267	.353*
Monitor execution of IT decisions	Pearson Correlation	.813**	1	.166	.162	.364*
Communicate IT issues	Pearson Correlation	.420*	.166	1	.185	.467**
IT governance framework	Pearson Correlation	.267	.162	.185	1	-.059
IT governance performance	Pearson Correlation	.353*	.364*	.467**	-.059	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

performance (p-value of less than 0.05). Details of inferential statistics are shown in Figure 6.

Therefore, the following hypotheses pertaining to formal governance mechanisms and governance performance are not rejected.

H2a: Organisations with formal governance co-ordination of decision-making mechanisms have high governance performance.

H2b: Organisations with formal governance execution monitoring mechanisms have high governance performance.

H2c: Organisations with formal governance communication mechanisms have high governance performance.

No significant correlation exists between awareness and usage frequency of various industry frameworks with IT governance performance. Therefore, the following hypothesis is rejected.

H2d: Awareness of industry frameworks related to IT governance is positively correlated with governance performance level.

Relationship between Organisational Context and Governance Performance

Results from the ANOVA test show that, with governance performance as dependent variable, the p-values for annual revenue (AR) and strength of workforce (SR) are more than 0.05. This indicates that there is no significant difference in governance performance across companies with different annual revenue and strength of workforce. However, effect size and power are both low, implying that there could be some differences in governance performance if a larger sample size is used. Details of inferential statistics are shown in Figure 7.

The following hypotheses are not rejected.

H3a: IT governance performance is not affected by annual revenue (organisation size).

H3b: IT governance performance is not affected by strength of workforce (organisation size).

As for primary goal (PG), the p-value is less than 0.05 with effect size of more than 0.14 and moderate power. Based on Cohen's (1984) criteria, effect size of more than 0.14 is considered important. This indicates there are some differences in governance performance between companies with primary goals of profit and growth. Further analysis revealed that governance performance for companies with growth as the primary goal is higher than that of profit.

Figure 7. ANOVA table for organisational context with governance performance as dependent variable

Tests of Between-Subjects Effects

Dependent Variable:IT governance performance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	1.757[a]	6	.293	1.769	.145	.290	10.615	.557
Intercept	258.656	1	258.656	1562.578	.000	.984	1562.578	1.000
AR	.194	2	.097	.587	.563	.043	1.174	.137
SW	.033	1	.033	.197	.660	.008	.197	.071
PG	.815	1	.815	4.924	.035	.159	4.924	.570
AR * SW	.380	1	.380	2.296	.142	.081	2.296	.309
AR * PG	.000	0	.	.	.	.000	.000	.
SW * PG	.000	0	.	.	.	.000	.000	.
AR * SW * PG	.000	0	.	.	.	.000	.000	.
Error	4.304	26	.166					
Total	392.799	33						
Corrected Total	6.061	32						

a. R Squared = .290 (Adjusted R Squared = .126)

b. Computed using alpha = .05

Therefore, the following hypothesis is rejected.

H3c: Organisations with profit as the primary goal have a high IT governance performance compared to organisations with growth as the primary goal.

DISCUSSION AND CONCLUSION

Research Findings

Ownership of Decision-Making

Research data show that IT governance performance is related to who makes the IT decisions. The right decision owner(s) making appropriate decision types lead to better IT governance performance. Specifically, it is recommended that:

1. Business and IT jointly make decisions on IT principles.
2. IT makes decisions on IT architecture, independently from business.
3. IT makes decisions on IT infrastructure, independently from business.
4. Business and IT jointly make decisions on business application needs.
5. Business makes decisions on IT investment, independently from IT.

Recommendations #1 to #4 are consistent with the pattern of top governance performers as recommended by Weill and Ross (2004). The slight divergence in the last recommendation will be discussed shortly (the reference study ranks this arrangement not as the best performer, but 2nd and 3rd best performers).

Once defined, IT principles require periodic review to align with changing corporate strategies and new possibilities arising from technology development. Joint ownership of business and IT in the formulation and review of IT principles ensures a balanced perspective of corporate strategy and technology development.

IT architecture and infrastructure are technical decisions based upon requirements articulated as part of IT principles. IT specialists are in the best position to make effective decisions based on technologies available in the market. It is important that IT specialists do not succumb to specific wants (not needs) of "IT-savvy and power users" of the organisation. Instead, they should make objective judgments in the interest of the organisation.

Fulfilment of specific business needs is the most obvious business value of IT. This explains why this is also the area where business exerts great pressure on IT to deliver value. It is important that IT actively participates in decision-making of business application needs. Creativity and discipline, though conflicting in nature, are important to guide IT when to turn down specific requests for functionality that deviates from architectural principles. What seems to be a short-term loss to the organisation will prove itself as the right decision in the long run. Compliance with architectural guidelines ensures integrity, flexibility and scalability to meet future business needs.

In almost all organisations, IT remains as a service centre with budget allocated from business. Investment and prioritisation decisions are best made by the business with input from IT. The only exception is technology renewal investments which are funded through periodic budget allocation from business, leaving the individual spending under discretion of IT executives/managers. The empirical results show that organisations are not ready to share investment decision-making powers with IT. Having businesses to make investment decisions will avoid potential problems of IT being caught in the middle when business units/processes dispute over the priority of IT projects, however it implies that IT has not gain the respect to be considered an equal decision-making partner.

Formal Governance Mechanisms

In deriving the hypotheses relating formal governance mechanisms to high governance performance, it was argued that good IT governance does not happen by chance. Coordination of decision-making process, execution monitoring and communication are the three categories of governance mechanisms.

Whether IT governance mechanism is viewed from a generic perspective, or based on specific COBIT control objectives, findings from this research showed that formal governance mechanisms are necessary to achieve good governance.

Contrary to the expected outcome, awareness of industry frameworks was found not to be related to IT governance performance. The level of awareness of IT governance was also not encouraging. Out of the 33 respondents, 12 (36.4%), 10 (30.3%) and 6 (18.2%) have not heard of COBIT, CMMI and ITIL respectively. COBIT and CMMI are currently used in only 5 (15.2%) organisations, while ITIL trailed closely with 4 (12.1%) organisations. Obviously more initiatives in this aspect are imperative.

On a positive note, when asked to describe what the term "IT governance" means, a significant number of respondents could accurately describe the issues underlying and goals of IT governance.

3 respondents stated precisely the fundamental concerns of IT governance: IT's delivery of value to the business and mitigation of IT risks. In their own words:

- *"IT governance means to measure, monitor and optimize the business value from investment in IT."*
- *"Management of IT performance and risk"*
- *"Compliance of IT department to rules and regulations set forth to ensure companywide risks and costs are minimized."*

Comments from several other respondents describe one or more objectives of IT governance, in terms of strategic alignment, value delivery, risk management, resource management and performance management. Quoted directly from the survey responses:

- *"Centralized Controlling for IT Cost, Headcount, and investment."*
- *"Control of flexibility of expanding"*
- *"Management of IT investment, service delivery, change of projects and etc"*
- *"Ensuring security and meeting requirements or compliance necessary for IT system within the corporation"*
- *"With the IT governance, It can bring the tool and knowledge used to benchmark and measure IT function to meet those needs with confidence and asurance."*
- *"Control and policies of IT & implementation"*

Organisational Context

The research findings show that organisational size is not a predictor of IT governance performance. This is consistent with the notion that IT governance affects big and small, profit and non-profit organisations.

However, it must also be noted that the distribution of annual revenue is skewed towards large corporations with significant representations of small and medium-sized companies. In addition, the distribution of strength of workforce is also skewed towards medium-sized companies and large corporations having more than 100 and 1000 employees respectively.

Furthermore, a surprising finding was made pertaining to influence of profit/growth-orientation towards IT governance performance. The research initially predicted that organisations with profit as the primary goal would implement standardised and tightly-controlled processes and tools, leading to higher IT governance performance. However, the research findings showed otherwise. Organisations with growth as the primary goal showed marginally higher IT governance performance. A bigger sample size is necessary to make further conclusions.

Governance Performance

Organisations in this research achieve a mean IT governance performance score of 3.42 with a standard deviation of 0.44 on a scale of 1 to 5. The lowest score is 3.00 while the highest is 4.56. This seems to be contradictory to the level of awareness concerning industry frameworks related to IT governance. The score was calculated independently, based only upon how important the outcomes are to the respondents and how successful IT governance has contributed to each outcome. A conflicting result indicates that the understanding and implementation of IT governance is on track, though familiarity with technicalities of the frameworks should be improved.

Limitations of Research

There are two major challenges faced throughout the research. Firstly, availability of data was a big hurdle. A number of organisations explicitly declined to participate in the survey, even though anonymity was assured. The number of responses received is sufficient to provide a preliminary view of the as-is situation of IT governance, but could be challenged statistically.

Secondly, IT governance itself is a relatively new topic in Malaysia. Some respondents reported confusion with usage of certain terms (e.g. corporate executives vs. business unit managers). Such clarity could only be achieved using case study research methodology with interviews and investigation of archival records in multiple organisations.

Generalisablility of Research Findings

This research was targeted at the electronics manufacturing industry in Malaysia, as the nation's leading industrial sector and having pervasive usage of IT in manufacturing process.

It is possible to generalise the research findings and empirical results to other industries. Past studies by Olson and Chervany (1980), Ahituv, Neumann and Zviran (1989), Clark (1992) found no relationship between industry and IT governance. In addition, industry type is found not to be a strong predictor in the study by Brown and Magill (1994).

Managerial Implications

The reality is that IT governance is neither a technical issue confined to the IT organisation, nor is the IT organisation solely responsible for success of IT service. Empirical results demonstrating that strong IT governance result in successful business outcomes have stood the test of time.

The best practices of who should make what IT decisions have been established through this research. Formal governance mechanisms have been shown to be necessary for good governance. IT governance does not just happen by chance or as an afterthought.

REFERENCES

Ahituv, N., Neumann, S., & Zviran, M. (1989). Factors affecting the policy for distributing computing resources. *Management Information Systems Quarterly*, *13*(4), 389–401. doi:10.2307/248722

Boynton, A. C., & Zmud, R. W. (1987). Information technology planning in the 1990's: Directions for practice and research. *Management Information Systems Quarterly*, *11*(1), 58–72. doi:10.2307/248826

Broadbent, M., & Kitzis, E. (2004). *The new CIO leader: Setting the agenda and delivering results*. Boston, MA: Harvard Business School Press.

Brown, C. V. (1999). Horizontal mechanisms under differing IS organization contexts. *Management Information Systems Quarterly*, *23*(3), 421–454. doi:10.2307/249470

Brown, C. V., & Magill, S. L. (1994). Alignment of the IS function with the enterprise: Toward a model of antecedents. *Management Information Systems Quarterly*, *18*(4), 371–403. doi:10.2307/249521

Clark, T. D. J. (1992). Corporate systems management: An overview and research Pprspective. *Communications of the ACM*, *35*(2), 61–86. doi:10.1145/129630.129633

Cohen, J. (1984). *Statistical power analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davenport, T. H., Hammer, M., & Metsisto, T. J. (1989). How executives can shape their company's information systems. *Harvard Business Review*, *67*(2), 130–134.

Dixon, P. J., & John, D. A. (1989). Technology Issues Facing Corporate Management in the 1990s. *Management Information Systems Quarterly*, *13*(3), 246–255. doi:10.2307/248998

Ein-Dor, P., & Segev, E. (1978). Organizational context and the success of management information systems. *Management Science*, *24*(10), 1064–1078. doi:10.1287/mnsc.24.10.1064

Ein-Dor, P., & Segev, E. (1982). Organizational context and MIS structure: Some empirical evidence. *Management Information Systems Quarterly*, *6*(3), 55–69. doi:10.2307/248656

Federation of Malaysian Manufacturers. (2006). About FMM. Retrieved September 11, 2006, from http://www.fmm.org.my/p_ho.asp?ThemeID=254

Garrity, J. (1963). Top management and computer profits. *Harvard Business Review*, *41*(4), 6–13.

Grembergen, W. V., & De Haes, S. (2005). Measuring and improving IT governance through the balanced scorecard. *Information Systems Control Journal*, *2*, 35–42.

Henderson, J. C., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *32*(1). doi:10.1147/sj.382.0472

Irani, Z., & Love, P. (2002). Developing a frame of reference for ex-ante IT/IS investment evaluation. *European Journal of Information Systems*, *11*, 74–82. doi:10.1057/palgrave/ejis/3000411

IT Governance Institute. (2003). *Board briefing on IT governance*. Rolling Meadows, IL: IT Governance Institute.

IT Governance Institute. (2007). *COBIT 4.1*. Rolling Meadows, IL: IT Governance Institute.

Loh, L., & Venkatraman, N. (1992). Diffusion of information technology outsourcing: Influence sources and the Kodak Effect. *Information Systems Research*, *3*(4), 334–358. doi:10.1287/isre.3.4.334

Malaysian Industrial Development Authority. (2006). Malaysia's electrical and electronics industry. Retrieved September 11, 2006, from http://www.mida.gov.my

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Office of Government Commerce. (2007). *The introduction to the ITIL service lifecycle book*. Norwich, UK: The Stationery Office.

Olson, M. H., & Chervany, N. L. (1980). The relationship between organizational characteristics and the structure of the information services function. *Management Information Systems Quarterly*, *4*(2), 57–69. doi:10.2307/249337

Rockart, J. F. (1988). The line takes the leadership. *MIT Sloan Management Review*, *29*(Summer), 57–64.

Ross, J. W., & Beath, C. M. (2002). New approaches to IT Investment. *MIT Sloan Management Review, Winter 2002*, 51-59.

Sambamurthy, V., & Zmud, R. W. (1999). Arrangement for information technology governance: A theory of multiple contingencies. *Management Information Systems Quarterly*, *23*(2), 261–290. doi:10.2307/249754

Software Engineering Institute. (2005). *Capability maturity model integration (CMMI) overview*. Pittsburgh, PA: Carnegie Mellon University.

Sohal, A. S., & Fitzpatrick, P. (2002). IT governance and management in large Australian organizations. *International Journal of Production Economics*, *75*(1-2), 97–112. doi:10.1016/S0925-5273(01)00184-0

Tan, K. S., Eze, U. C., & Teo, W. L. (2008). Information technology governance in the Malaysian electronics manufacturing industry. *Proceedings of 10th Innovation and Knowledge Management in Business Globalization: Theory & Practice* (pp. 587-593), Kuala Lumpur:IBIMA.

Tanriverdi, H. (2006). Performance effects of information technology synergies in multibusiness firms. *Management Information Systems Quarterly*, *30*(1), 57–77.

Tavakolian, H. (1989). Linking the information technology structure with organizational competitive strategy: A survey. *Management Information Systems Quarterly*, *13*(3), 308–318. doi:10.2307/249006

Van Grembergen, W. (2002). Introduction to the minitrack IT governance and its mechanisms. *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS)*, (pp. 7-10), Island of Hawaii:IEEE.

Van Grembergen, W. (2003). *Strategies for information technology governance*. Hershey, PA: Idea Group Publishing.

Van Grembergen, W., & Van Bruggen, R. (1997). *Measuring and improving corporate information technology through the balanced scorecard technique.* Paper presented at the Proceedings of the Fourth European Conference on the Evaluation of Information Technology, Delft, The Netherlands.

Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior results*. Boston, MA: Harvard Business School Press.

Weill, P., Subramani, M., & Broadbent, M. (2002). Building IT infrastructure for strategic agility. *MIT Sloan Management Review*, *44*(1), 57–65.

This work was previously published in Enterprise IT Governance, Business Value and Performance Measurement, edited by Nan Si Shi and Gilbert Silvius, pp. 41-60, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 50
An Environmentally Integrated Manufacturing Analysis Combined with Waste Management in a Car Battery Manufacturing Plant

Suat Kasap
Hacettepe University, Turkey

Sibel Uludag Demirer
Villanova University, USA

Sedef Ergün
Drogsan Pharmaceuticals, Turkey

ABSTRACT

This chapter presents an environmentally integrated manufacturing system analysis for companies looking for the benefits of environmental management in achieving high productivity levels. When the relationship between environmental costs and manufacturing decisions is examined, it can be seen that the productivity of the company can be increased by using an environmentally integrated manufacturing system analysis methodology. Therefore, such a methodology is presented and the roadmap for generating environmentally friendly and economically favorable alternative waste management solutions is elaborated. The methodology combines data collection, operational analysis of the manufacturing processes, identification of wastes, and evaluation of waste reduction alternatives. The presented methodology is examined in a car battery manufacturing plant, which generates hazardous wastes composed of lead. It is aimed to decrease the wastes derived from the production so that the efficiency in raw materials usage is increased and the need for recycling the hazardous wastes is decreased.

DOI: 10.4018/978-1-4666-1945-6.ch050

INTRODUCTION

Manufacturing companies are now aware that whether to implement sustainable environment and resource management practices or not is no longer a choice for them. Companies operate in a world of dynamic competition in which technology, production and manufacturing processes, customer needs and, environmental regulations are constantly changing. Therefore, companies should constantly find innovative solutions to survive under the pressure of competitors and regulators. The increasingly strict environmental regulations combined with the improving consciousness of consumers for environmentally friendly products have put manufacturers in a precarious situation. Consequently, not only researchers, but also manufacturing managers are recognizing the importance of environmental management systems used for managing environmental practices (Angell and Klassen, 1999; Claver et al., 2007; Gupta and Sharma, 1996; Klassen, 2000; Porter and van der Linde, 1995; Sroufe, 2003; Xigang and Zhaoling, 2000).

A major barrier to the adoption of environmental management systems is that companies often do not know the environmental costs of operating their business and therefore do not know the financial benefits that can be obtained by reducing their environmental impacts. Previously, environmental costs were generally defined as costs dealing with environmental laws, regulations, and taxes. It is now recognized that the true environmental costs includes: costs of resources, waste treatment and disposal costs, the cost of poor environmental reputation, and the cost of paying an environmental risk premium. The calculation and evaluation of environmental costs provide better understanding of the production cost of a product, and that it properly allocates costs to product, process, system or facility. Measuring environmental costs also improves the correctness of the pricing and gives profitability and competitive advantage, therefore increases the overall management system of a company. For this reason, environmentally integrated manufacturing decisions require for the consideration of technical, economic, and ecological aspects of the manufacturing processes simultaneously. Especially the companies using hazardous materials in their production have started to consider environmentally integrated manufacturing systems to decrease their impacts on environment and to prevent pollution at source directly.

This chapter presents a methodology for the environmentally integrated manufacturing system analysis for companies aiming to achieve the benefits of environmental management in obtaining high productivity levels. The aim of the methodology is the reduction of wastes derived from the manufacturing processes. The wastes produced during the manufacturing process are important cost issues for manufacturers. Waste of raw material creates important costs and environmental effects, especially when it is hazardous. For the case of using hazardous raw materials, the wastes derived from the manufacturing processes can not be sent to trash, instead they are sent to the recycling or treatment facility. Obviously, wastes that can not be used within the facility create inefficiency in the usage of raw materials. Therefore, the problem considered is to decrease the formation of wastes in order to decrease waste management costs and improve raw material usage. One solution to this problem is to use an environmentally integrated manufacturing system point of view aiming to decrease product losses, while reducing costs and improving profitability. The presented methodology is examined in a car battery manufacturing plant since any improvement to reduce wastes in this company will be yielding immediate environmental benefits since the waste usually consists of lead, which is hazardous. This application forms also an example about the achievability of "cleaner" production philosophies in the manufacturing sector using hazardous raw materials. The presented methodology is designed in a way that its implementation in different sectors is also possible. The presented methodology is the combination of data collection, manufacturing system analysis, identification of wastes, and evaluation of alternative waste management solutions. The objectives of the presented methodology are

- to create a point of view about the types of alternative waste management solutions that can be determined and to show their benefits,
- to discuss how the operations management functions of the facility will be affected by the implementation of an alternative waste management solution,
- to discuss what types of limitations can be encountered in a facility while implementing an alternative waste management solution,
- to give an example about the practicability of pollution prevention philosophies to the manufacturing sector.

The presented methodology intends the integration of environmental management principles with the operations management. In the next section, the need for adopting a sustainable development perspective is made clear by analyzing the relationship between environmental costs in manufacturing decisions. In addition, environmental operations management principles and criteria affecting the environmental operations management strategy are described. Next to this section, the methodology for the environmentally integrated manufacturing system analysis is presented. After description of the methodology, the manufacturing system of a car battery manufacturing plant is analyzed by using the given methodology. Finally, the conclusion and summary of the chapter is given.

BACKGROUND

In the last few decades environmental problems have received increasing attention. Protection of the environment has become an issue at all levels of society. Within the field of operations management, attention for environmental issues is now growing rapidly. Operations management is the process of managing people, resources and production systems in order to convert inputs into outputs. The inputs of the system are energy, materials, labor, and capital. The outputs of the system are the products demanded by customers (Nahmias, 2004). Wastes can be produced when converting raw materials into products and are considered as non-value-added outputs that create extra costs to the production of the product. The reduction or elimination of wastes in the production has always been a goal in operations management e.g., lean manufacturing. In order to provide the best use of raw materials, operations managers have to minimize the amount of the waste produced, namely, to reduce the cost and impact of the waste. An efficient approach undertaking this target is to integrate waste management into the operations management decision. Operations management deals with the optimization of the manufacturing process and covers decisions involving production planning, scheduling, capacity planning, inventory management, material management, workforce management, and quality management. It can be remembered that the traditional objective of inventory control, scheduling and production planning is to minimize production costs and improving measures such as worker idle time, work-in-process inventory levels, and production lead times while recognizing constraints such as limited space. Undoubtedly, by the integration of waste management into the operations management, the objective function should include not only the cited objectives but also environmental ones. It is obvious that an operations management team is not only responsible of the achievement of the desired products in terms of quality and quantity, but also has to control working practices, resources consumption, emissions, and the flow of hazardous materials. Thus, operations managers are directly concerned with environmental issues in their operational responsibilities and play a critical role in developing management systems that affect environmental performance (Angell and Klassen, 1999; Gupta and Sharma, 1996).

Environmental Operations Management (EOM) is a concept of integrating environmental management principles into the operations management process for the conversion of resources into usable products. Due to EOM principles, instead of looking at environmental management as a "cost", companies can use EOM as an opportunity to improve their position by eliminating waste, removing non-value added materials and equipment, and reducing both short-term cost and long-term liability. The company will be able to plan and design products that do not create toxic wastes or require environmentally hazardous processes – from the introduction of the product to its final disposition. A long-term competitive advantage can be obtained by setting EOM principles to the production process. In order to create an environmental operations strategy, the operations management team is constrained by criteria such as: dependability, efficiency, flexibility and quality. The quality of the environmental attributes of the products is important criteria for consumers willing to buy. The dependability of a company is affected by the use of hazardous materials and processes. Accidents involving the shipment, transfer, and use of hazardous materials often result in temporary shutdowns of manufacturing plants. In addition, the capital investment needed to process and dispose hazardous materials is often high and affects the dependability. The flexibility of the company is limited by the materials and processes necessary for the production and by the types and quantities of hazardous materials discharged into the environment. Efficiency can be achieved by finding less expensive and less environmentally hazardous materials and processes to manufacture the desired products (Gupta and Sharma, 1996).

As mentioned before, one major barrier to the adoption of environmental management systems is that companies often do not know the environmental costs of operating their business and therefore do not know the financial benefits that can be obtained by reducing their environmental impacts. Until only a few decades ago, there was a belief that any investment in improved environmental performance would contribute to increased costs, which will finally reduce profits. Previously, environmental costs were generally defined as costs dealing with environmental laws, regulations, and taxes. Firms have tended not to measure environmental costs because management accounting systems have focused on clearly identifiable costs but not on the costs and benefits of alternative actions. In 1991, Porter put forward a new standpoint to the interaction between profitability and pollution prevention (Porter, 1991). This interaction has increased the theoretical and practical interest in the possibility that profitability and pollution reduction were not conflicting goals. According to Porter, pollution was simply a diminishing value in the production and was an indication of problems in products and/or processes. Therefore, contrary to previous opinions, reducing or eliminating pollution would not weaken but strengthen corporate competitiveness. After Porter's study, a radical change has come in management's views on pollution reduction and better environmental management. Companies became aware of the important role that environmental costs play in the calculation of total costs of production. It is now recognized that the true environmental costs includes: costs of resources, waste treatment and disposal costs, the cost of poor environmental reputation, and the cost of paying an environmental risk premium. With this point of view, environmental costs are transferred from overall (or indirect) costs to direct costs. As a result, the calculation and evaluation of environmental costs provide better understanding of the production cost of a product, and that it properly allocates costs to product, process, system or facility. Measuring environmental costs also improves the correctness of the pricing and gives profitability and competitive advantage, therefore increases the overall management system of a company. Thus, environmentally integrated manufacturing decisions require for the consideration of technical, economic and ecological aspects simultaneously (Melnyk et al., 2003; Spengler et al., 1998; Wu and Chang, 2004). Additionally, companies try to avoid inefficiencies in the production and waste removal charges as a result to decrease the cost of unit production. Especially the companies using hazardous materials in their production have started to consider environmentally integrated manufacturing systems to decrease their impacts on environment and to prevent pollution at source directly.

Manufacturing decisions should not be made in isolation from decisions in environmental management. In recent years, many production planners and decision makers have started to recognize the role and significance of the integration of economic and environmental efforts in a single production-planning program. New concepts connecting manufacturing practices, pollution control and prevention and operations are recently being used in order to increase the efficiency of converting raw materials into products. Specifically, studies are carried out in order to create decision support tools for analyzing the effects of planning decisions on the amount of product losses. Companies are willing to organize their production systems to enhance resource productivity by adopting an environmental approach. Based on this kind of approach, a company can include environmental principles in the mission statement, incorporate the cleaner production philosophy into product and process design and, consequently, develop an environmental business strategy in order to gain competitive advantage. At this point, it is important to understand the difference between pollution prevention and pollution control. Pollution control covers the elimination of pollution after the waste is generated. On the other hand, pollution prevention covers the modification/redesign of the production process and the introduction of new technologies throughout the product life-cycle to identify the source of the problem. A company adopting traditional pollution control methods focuses on its activity on the short term. In this context, the company sets as its main aim to carry out environmental impact correcting actions that do not entail the development of new skills needed to manage new environmental processes. Therefore, it is seen that traditional pollution control methods are practically inefficient compared to prevention methods. Therefore it is clear that preventing environmental damage is cheaper and more effective than attempting to manage or fix it. Pollution prevention has replaced the traditional pollution control methods and has become an important research topic for the process design (Claver et al., 2007; Xigang and Zhaoling, 2000). There are many studies demonstrating that pollution prevention is almost the most cost-effective constituent of integrated waste management strategies in different manufacturing systems (Akkerman and van Donk, 2006; Dahab et al., 1994).

One step forward of pollution prevention is the concept of cleaner production (CP). CP can be described as the continuous application of an integrated preventative environmental strategy to processes, products and services to increase efficiency and reduce risks to humans and the environment. For production processes, CP includes conserving raw materials and energy, eliminating toxic raw materials, and reducing the quantity and toxicity of all emissions and wastes before they leave a process. For products, the CP strategy focuses on reducing impacts along the entire life cycle of the product, from raw materials extraction to ultimate disposal of the product. In CP applications, savings are often achieved with little or no capital expenditure by simply changing management practices. Many successful case studies show that CP can provide opportunities for making sound choices for both environmental concerns and economic benefits (Dahodwalla and Heart, 2000; Jia et al., 2005). One of the most basic reputations of CP is that it improves efficiency and productivity in industry. These improvements result in lower expenditure on resources such as energy and water, increased efficiency in production, fewer risks associated with environmental impacts, and decreased waste generation that leads to savings in landfill fees and pollution licenses. Due to the CP context, continuous improvement is required not only in technology and know-how, but also in managerial skills and policies. CP offers much for operations management researchers to draw on as they explore the linkages between process and product technology, environmental management and performance.

The car battery manufacturing industry is one of the critical industries in the implementation of pollution prevention and CP studies because of the lead (Pb) used in production. Wastes should be decreased in order to prevent the hazardous effects of Pb to the environment and human health. Many studies involving technology changes and pollution prevention options are accomplished for the battery manufacturing industry (Boden and Loosemore, 2007; Dahodwalla and Heart, 2000; Ferreira et. al. 2003).

THE ENVIRONMENTALLY INTEGRATED MANUFACTURING SYSTEM ANALYSIS

The environmentally integrated manufacturing system analysis provides a roadmap for companies willing to apply an integrated preventive waste management approach in their manufacturing plant. Many possible working areas and information to be collected are given so that this methodology can be applied in different manufacturing sectors and companies. The design of the methodology firstly involves the investigation of techniques aiming the reduction of wastes in the manufacturing area. Waste auditing methodologies are examined in this context. Operations management techniques involving the reduction of material losses are then analyzed. The problem solving techniques are used for the identification of manufacturing problems, the investigation of possible solutions and the selection of the best solution. The methodology is established by integrating operations management perspectives to the classical waste auditing methodology. Waste auditing generally covers three steps. The first step, which is the pre-assessment, covers the division of the processes into unit operations and the construction of the process flow diagram. The second step, which is the material balance, covers the identification of inputs, outputs and the current reuse and recycling methods. The final step which is the synthesis, deals with the evaluation of waste reduction options and the creation of the waste reduction action plan. The presented methodology, different from the classical waste auditing, cautiously focuses on the operations of the manufacturing facility. Processes are analyzed based on the inputs used, working practices, capacities of machines, and byproducts and outputs derived from the processes. Steps of the presented methodology can be summarized as follows:

- **Step 1:** Analysis of the system
    - Data collection
    - Manufacturing system analysis
    - Construction of a process flow diagram
    - Current waste management practices
    - Environmental impacts of materials used in production
- **Step 2:** Identification of Alternative Waste Management Solutions
    - Investigation of Waste Management Technologies
    - Investigation of Alternative Waste Management Solutions

- **Step 3:** Evaluation of the Alternative Waste Management Solutions
  - Operational Analysis of Alternative Waste Management Solutions
  - Technical Analysis of Alternative Waste Management Solutions
  - Economical Analysis of Alternative Waste Management Solutions
- **Step 4:** Comparison of Alternative Waste Management Solutions

Analysis of the System

In the first step of the methodology, the analysis of the manufacturing system should be accomplished. This analysis will provide the information about the current manufacturing and waste management system. The data collection phase involves a thorough assessment of the company, from the purchase of various inputs (e.g., raw materials, energy and water) to the output itself (products). This phase implicates the collection of data such as: product types, raw materials used and their prices, the demand of products and their prices, material requirement planning, and supply chain partners. By analyzing the data collected, it will be possible to find the processes operating with low efficiencies. The manufacturing process is considered as one of the important sources of environmental impacts of the industrial production. Therefore, it is essential to analyze in detail all the elements of the manufacturing system in order to prevent the pollution at the source. The information about the system as a whole is gathered in this phase by combining the data collected about processes and their operational conditions, material management, capacity planning, production planning and scheduling, quality management, workforce management, supply chain management, and logistics. The operations forming wastes are determined in this phase of the system analysis. It is important to obtain real data in this step, because even if it is not optimal, the current manufacturing system of the company will be a basis for a comparison. With the aid of data collected in the first and second phases, it is now possible to generate a process flow diagram for a complete manufacturing process including sequence of operations, raw materials, semi-finished products, products and wastes.

In the next phase, the investigation of current waste management practices in the company is accomplished. The waste management system of a company involves the entire procedure of collecting, transporting, processing, recycling or disposal of waste materials. Waste management can involve different methods and fields of expertise for solid, liquid, gaseous or radioactive substances, with the aim of reducing the effects of these wastes on the health and environment. During this phase, data such as the quantity and rate of generation of production wastes are collected. The content of wastes and the waste management costs are also investigated. Current waste management practices in the company should be determined correctly and then analyzed with respect to their appropriateness. A waste management application may be efficient in treating a waste completely, but it may be very costly or may cause extra usage of raw materials. On the other hand, a simple and inexpensive application may be financially welcomed but may gradually create serious damages in worker's health or environment. As a result, measurements should be made to find out the real needs of the system in order to solve each problem efficiently. For the last phase of the first step, investigating the environmental impacts of raw materials, production processes and products is important. Investigations may give ideas about the substitution of hazardous raw materials with non-hazardous materials. These kinds of substitutions may give benefits such as the decrease in employee exposure to pollution and decrease in costs due to the hazardousness of the material, such as special transportation costs, fees, etc. Making detailed analysis about the environmental impacts of materials is also an important action for convincing people to implement the alternative waste management solutions suggested. The workers and the management of the company can be influenced to take environmental precautions by informing them about impacts of products produced or materials used. After clarifying the environmental impacts of raw materials, production processes and products, it will be effective to suggest alternative waste management solutions as a new way of achieving better working conditions in the facility.

Identification of Alternative Waste Management Solutions

In the second step of the methodology, the investigation of possible alternatives aiming to reduce or reuse waste should be performed. Once the production wastes and inefficient processes in manufacturing system are identified, the next move is to investigate the available waste management technologies or practices in the literature. Current operational conditions of the company must be compared with the conditions reported in the relevant literature. The investigations may draw upon basic research results, literature searches, field research, and discussions with industry experts and technology-users. This process will make it possible to recognize environmentally friendly methods, the best available technologies, and the newest technologies. Alternative waste management solutions are prepared by using information about wastes and processes operating with low efficiencies. Alternative waste management solutions involving the integration of economic and environmental efforts are investigated. According to this integrated point of view, waste is accepted as a resource that can be used in the production process. The investigation of alternative waste management solutions will pursue the following hierarchy:

1. Reduction at source
2. Internal (on-site) and external (off-site) recycling
3. Treatment of waste streams
4. Controlled disposal when there is no other solution.

The reduction of wastes at source may be accomplished by the purchase of new equipments or materials, or may simply involve small changes in the production process. Pollution prevention alternatives aiming the reduction of waste at the source are at the first priority. The environmental and economic benefits of reducing waste at source are much higher than other alternatives because the raw material usage and the waste management costs can both be reduced. If the formation of waste can not be prevented at source, internal and external recycling alternatives must be considered in order to minimize the waste treated/disposed. With the recycling opportunities, wastes from one industrial process can serve as a raw material for another, therefore the impact of the industry on the environment is reduced. Specifically, on-site recycling is very valuable because it eliminates the cost of sending the waste to an off-site facility, therefore increases the productivity of the company. In cases where the waste can not be prevented or can not be reused within the facility, off-site recycling is a choice allowing to send out the waste from the facility and to allow its reuse by other industrial processes. Therefore, the impact of the waste to the environment is diminished. However, off-site recycling creates a cost due to the transportation of wastes to the recycling facility, and more importantly, due to the inefficient use of raw material that can not be converted into a product.

The treatment of waste streams is accomplished to remove the hazardous portion of a waste from the non-hazardous portion, such as water. Various techniques are available to reduce the volume of a waste through physical treatment. For example, concentration techniques including vacuum filtration, filter press or heat drying are commonly used to dewater the sludge. These techniques are used to reduce the volume, and thus the cost of disposing a waste material. Moreover, once the material is concentrated, there is a greater likelihood that the materials in the waste can be recovered. This allows the potential use of the waste streams as a raw material for other companies or for the company itself (Freeman, 1988, p. 5.13). Controlled disposal is a solution for wastes which can not be treated by pollution control technologies. Disposal is the last preferred waste management option. Nevertheless controlled disposal is an important part of environmental management; even though it is the least effective one.

Evaluation of Alternative Waste Management Solutions

Each investigated alternative waste management solution in the previous steps of the methodology must be analyzed in order to evaluate how beneficial or practical the implementation of this alternative will be. It is important to use an interdisciplinary point of view while evaluating each alternative waste management solution since the entire production system is aimed to be optimized. By this integrated point of view, it will be possible detect the positive and negative results of each alternative waste management solution proposed. The scope and complexity of an apparently feasible alternative waste management solution can change after the analysis of the initial problems or after the design of the system. An alternative waste management solution is feasible if it works with available or obtainable resources, produces better results in both environmental and economical point of view and if it does not conflict with other management functions. For example, while an alternative waste management solution may be feasible due to machine schedules, it may be infeasible due to the demand pattern of the products. As a result, it may not be possible to implement this alternative. It is also important for an alternative waste management solution to have a reasonable reimbursement time. Evaluations of solutions alternatives are made under the title of the following measures.

Operational Analysis of Alternative Waste Management Solutions: Each alternative waste management solution should be analyzed in order to see if it is operationally feasible. The operational viability generally consists of measures such as the applicability of an alternative. In order to see if an alternative waste management solution is desirable in an operational sense, it should be proved that the alternative is practical and efficient. In other words, it should be demonstrated that the alternative makes maximum use of available resources including raw materials, people, and time. Before evaluating an alternative waste management solution, it should be verified if the current work practices and procedures support a new system. This verification also covers the adaptation of end users and managers to the change. The flexibility and expandability of the alternative must also be taken into account because capacity changes can take place in the company as a result of future needs and projected growth.

Technical Analysis of Alternative Waste Management Solutions: Each alternative waste management solution should be analyzed in order to see if it is technically feasible. The technical viability generally consists of measures such as the practicality of an alternative and the availability of technical resources and expertise within the facility. An important aspect identifying if the alternative waste management solution is technically feasible is the availability of the required technology. Some alternatives may involve solutions that can be applied without the need of a new technology or equipment. Conversely, there may be alternative waste management solutions that need the investigation of the market availability of the required technology or equipment. If the required technology is produced or sold within the country, it is needed to compare the different sellers and selling prices. On the other hand, if the required technology is exported, it is needed to investigate the different countries, selling prices and transshipment rates. In order to see if an alternative waste management solution is desirable in a technical sense, the company should verify if the technical requirements, impediments and competing technologies are identified. By evaluating current technology options and limitations, the company can define how difficult it will be to build the new system. For some alternative waste management solutions, there may be critical elements that require feasibility demonstration. Therefore the company may require making tests and experiments in order to see if the considered alternative waste management solution is technically feasible. By this way, the performance of experiments and tests can be compared with preliminary technical requirements and objectives. The experiments and tests will enable the presentation of a path forward for the next stage of the alternative. Moreover, written results, models, or laboratory process outputs demonstrating technical concepts and benefits will improve confidence that the alternative waste management solution will successfully meet the goals of the company.

Economical Analysis of Alternative Waste Management Solutions: If an alternative waste management solution is operationally and technically feasible, then the economical analysis of the alternative should be carried out to examine if it is affordable or not. The economical viability generally consists of measures such as the quantitative estimation of benefits of an alternative waste management solution, which are typically reduction in costs or risks. An alternative waste management solution that is economically desirable is regarded as justified because the economic aspects are generally the bottom line of many projects. In the economical analysis phase, the costs, benefits and incomes of the current system should also be identified. By this way, it will be possible to compare the economic aspects of the current system and the suggested alternative waste management solution. This will allow realizing the cost of not developing the new system. In order to see if an alternative waste management solution is desirable from an economical point of view, the cost-benefit analysis should be made and the cost-effectiveness of the alternative should be proved. By this analysis, it will be demonstrated if the benefits outweigh the estimated costs of development, installation, operation and maintenance. When making an economical analysis, it is important to predict tangible and intangible benefits. Furthermore, the timing of costs and benefits is an important factor determining the payback period, which is also an important decision criterion affecting the applicability of an alternative waste management solution.

Comparison of Alternative Waste Management Solutions

As it is known from operations research, optimizing a variable may not always give the overall optimal solution, especially when several aspects such as productivity and environmental performance are both aimed (Nahmias, 2004, p. 671). Therefore, the consequences of the alternative waste management solutions and their effects on the overall performance should be discussed. If an alternative waste management solution is found to be feasible, the company may decide to proceed with this solution after comparing it with other alternatives generated for the considered problem. In selecting the best alternative, a company often considers trade-offs. The final decision can be determined by working with end-users, reviewing operational, technical and economic data. On the other hand, decisions criteria such as dependability, efficiency, flexibility and quality may be used for creating an environmental operations strategy. Therefore, if a company aims to accomplish an environmentally integrated manufacturing system analysis, a multi-disciplinary team should be formed and operational, technical and economic aspects should be compared for each alternative waste management solution by taking account the criteria mentioned above. It is also essential that the company reviews and changes its environmentally integrated manufacturing plans dynamically to assure high environmental standards. An important issue in improving the manufacturing system is the use of feedback information in order to continually adjust the mix of inputs and technology needed to achieve desired outputs. Information derived from environmentally integrated manufacturing system analysis will continually change production planning decisions such as demand forecasting, purchasing, production and personnel scheduling, quality control, and inventory control issues. The team leader should constantly monitor the manufacturing system and its environment in order to plan, control and improve the system.

IMPLEMENTATION OF THE ENVIRONMENTALLY INTEGRATED MANUFACTURING SYSTEM ANALYSIS IN A CAR BATTERY MANUFACTURING PLANT

The implementation of the methodology is accomplished in a car battery manufacturing plant working with hazardous materials such as lead and sulfuric acid. The lead-acid battery manufacturing company occupies a total space of 22,500 m2 in its manufacturing plant in Ankara. The company produces different types of lead-acid batteries and also supplies different types of semi-finished products like grids, raw plates, charged plates as well as lead monoxide. The production capacity is 1,500,000 batteries per year and there are 50 white-collar and 250 blue-collar workers. The following assumptions are considered in the context of the study:

- The manufacturing process of wet-charged batteries is considered.
- Prices of products are assumed to be constant for every customer.
- Data collected about the manufacturing system is considered as input data.
- The unit waste management cost of an on-site recycled waste consists of its raw material cost. Labor and overhead costs associated to the considered waste are included if a change in these costs is obtained according to the evaluation of alternative waste management solutions.
- The unit waste management cost of an off-site recycled waste consists of its raw material cost and transportation cost, apart from the selling price to the recycling facility.
- The unit waste management cost of discarded wastes consists of the raw material cost.
- The unit waste management cost of wastewater consists of the raw material cost. Energy usage in the wastewater treatment facility is ignored. Chemical usage costs are included to costs calculations if a change in chemical usage is obtained according to the considered alternative waste management solution.
- In the net present value calculations, the costs are assumed to derive at the end of each month and the monthly interest rate is assumed to be 1.6%.
- The company works 6 days a week. Each day consists of 3 shifts, which are 8 hours each.
- The company does not hold end-product inventory and there is no cost associated to inventory carrying.
- There is no problem associated to the limited storage area.

With the use of the methodology, the manufacturing system of the company is analyzed and the alternative waste management solutions are investigated and evaluated.

Analysis of the System

The analysis of the system provides information about the current manufacturing and waste management system. The data collection phase involves a thorough assessment of the company, from the purchase of various inputs to the output. Data such as the types of batteries, demands and prices of each product, raw material prices and material requirement plans are obtained in this phase of the study. The demand forecasting procedure and bill of material of a car battery are obtained. The current production planning and scheduling system are analyzed. Each manufacturing process is analyzed in detail to determine the operations forming wastes. The manufacturing process of a car battery and wastes derived from each unit process are shown in Figure 1. The waste management system of the company involves the entire procedure of collecting, transporting, processing, recycling and disposal of waste materials. Information about the production wastes (quantity and rate of generation) is collected. The composition of wastes and waste management costs are investigated. Current waste management practices in the company are determined and analyzed with respect to their appropriateness in terms of environmental and economical benefits. Making detailed analysis about the environmental impacts of the wastes is also important for convincing people to implement the suggested alternative waste management solutions, which can be considered as a new way of achieving better working conditions in the facility. This study focuses on the wastes produced in three manufacturing processes as shown in Figure 1: grid casting, battery closing, wet-charging. Due to the current waste management system of the company, the identified wastes are considered as:

- On-site recycled wastes: Wastes used within the facility.
- Off-site recycled wastes: Wastes sent to recycling facilities to recover the raw material.
- Wastewater: Contaminated water treated in the wastewater treatment facility of the company

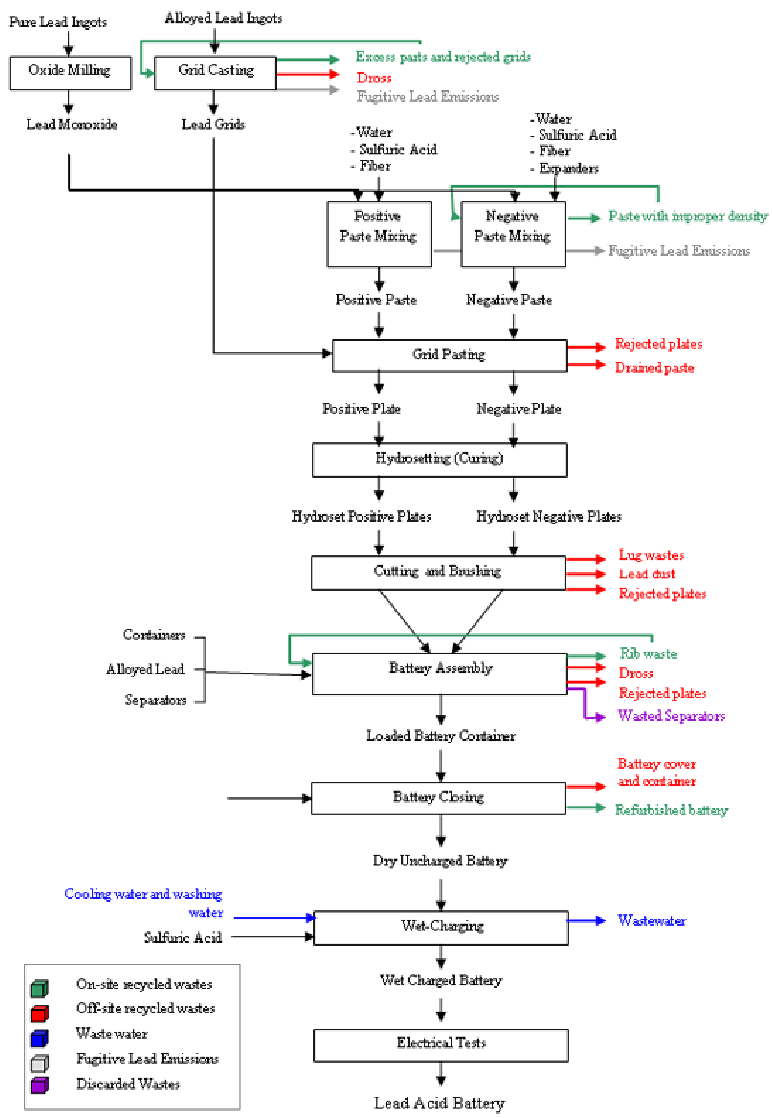
Identification of Alternative Waste Management Solutions

The alternative waste management solutions are identified by investigating waste management technologies and alternative waste management solutions. The possible waste management technologies in the battery manufacturing plant are identified and the benefits of implementing pollution prevention techniques in the considered sector are investigated. The best available technologies aiming the decrease of particular types of wastes are described in detail in the evaluation of alternative waste management solutions investigated for the considered wastes. Additionally, two case studies concerning a general pollution prevention study in the car battery manufacturing industry are examined. The applicability of each alternative waste management solution is evaluated in order to show the type of problems that could derive from their applications. By making an operational, technical and economical analysis for each of the alternatives, it is possible to present realistic propositions. Therefore, the possible benefits and limitations are identified for each solution alternative.

Alternative Waste Management Solutions: Wastes from Grid Casting

Dross obtained from the grid casting process is the waste that causes the highest lead loss in the company. In the grid casting process, the excess parts of each grid are cut automatically and transferred to the melting pot by a conveyor belt located behind grid casting machines. Additionally, the rejected grids detected by operators are added to the melting pot immediately after their detection. Therefore, the grid casting process involves continuous addition of wastes into the melting pot. The formation of dross is the result of molten lead alloy and air contact. It is simply the oxidation of lead alloy. Therefore, the formation of dross is inevitable on the surface of the molten

Figure 1. Process Flow Diagram of a Car Battery



lead alloy in the pot that is open to atmosphere. In fact, the dross formed may act as a protective passive film preventing further dross formation if it is not disturbed. But, when wastes deriving from the grid casting machine are added to the melting pot continuously, the surface covered by dross collapses, and this procedure causes an increase in contact area of lead alloy with air. As a result, the amount of dross increases continuously when wastes are added to the pot. There are two alternative practice identified for the reduction of dross amount in this study as explained and analyzed below.

Addition of Excess Parts and Rejected Grids in Batch Mode: The practice of adding excess parts and rejected grids may be changed from continuous to batch. In other words, the parts and grids may be collected and then added at specific time intervals, so that the contact time of the molten lead alloy with air can be decreased.

This practice is experimented in the grid casting procedure to observe any change of dross formation rate in this study.

Two experiments are accomplished for observing the relationship between dross formation and time interval adopted for the collection of rejected parts and grids. Two different time intervals were tested as 1.5 and 3.0 hour. The current grid casting procedure was also observed for comparison. Before the experiment, the dross layer from the surface of the molten lead alloy was completely removed from the melting pot by the use of a ladle with holes in it. The amount of dross formed in the melting pot during 1.5 and 3.0 hour periods was observed under the conditions in which the excess parts and rejected grids were collected. At the end of the collection period, the production of grids was stopped and the dross formed on the surface was removed and weighed. Following this, the collected excess parts and rejected grids were weighed and added to the melting pot. Since the excess parts and rejected grids weighed about 700 kg, the manual addition of the rejects to the pot took approximately 25 minutes. Then, the formation of dross for an hour was allowed under the conditions of no production. Finally, the dross formed in an hour was collected and weighed. The summation of the dross collected before and after the addition of rejects is considered as the total weight of dross obtained from the experiment with the considered time interval. The results from this experiment show that the ratio of rejects to the ratio of product produced is about 58-60%, which is considerably high.

Following up these steps, the current grid casting procedure was observed for the same time intervals; specifically 1.5 and 3 hours time periods. Accordingly, the production of grids was started with continuous feeding of the melting pot with the excess parts and rejected grids by conveyor belt. The amount of dross formed at the end of the time interval was measured by collecting and then weighing the dross. During this step, the amount of excess parts and rejected grids could not be measured since they are automatically added to the melting pot.

According to the results obtained from the two experiments with different time intervals, it is proved that the suggested work practice reduces the amount of dross significantly. It has also been shown that decreasing contact time decreased the amount of dross formed. Therefore, the longest possible time period for collection is expected to decrease the dross amount most. The cash flow analysis of the batch mode and the current system are compared in Table 1. The calculations in Table 1 take into basis that dross is reduced by 48% with the use of the batch mode alternative. The cost of raw material loss is calculated based on the unit cost of dross, which is 1.18 $/kg. Since it took 3 hours to obtain 41 kg of dross in the experiment, the production of dross is assumed to be 13.6 kg/hour. Additionally, the unit labor cost for the company is calculated to be 1.63 $/hour, based on the base wage rate. Given the unit labor cost and unit production rate of dross, it can be concluded that the unit direct labor cost of dross is 0.11 $/kg. Based on the monthly quantity of dross, the direct labor cost associated to dross in the current system is:

Direct Labor Cost of Current System= 0.11 $/kg x 12,299 kg/month = 1352 $/month.

In the production cost of a battery, it is known that the direct labor occupies a percentage of 14%, and the overhead occupies a percentage of 26% of the total cost. The overhead cost associated to dross can be calculated by using these ratios and the direct labor cost as calculated above. Knowing that the 14% of the total cost is 1,337 $/month, it can be calculated that the 26% of the total cost, which corresponds to the overhead cost associated to dross, is 2,712 $/month, as shown in Table 1. The direct labor and overhead costs of the batch mode are calculated similarly. In addition to the current system, in the batch mode there is a need of labor when collecting excess parts and rejected

Table 1. Comparison of cash flow analysis of the current system and the batch mode for t=3.0 hours

	Current System	**Batch Mode**
Quantity of dross (kg/month)	12,299	6,396
Cost of raw material loss ($/month)	14,605	7,593
Direct labor cost associated to dross ($/month)	1,352	759
Overhead cost associated to dross ($/month)	2,712	1,410
Labor cost associated to the collection of parts ($/month)	-	285
Net Present Value ($/year)	**203,552**	**108,920**

grids. In the experiment, it is seen that a total of 45 minutes is spent for a production of 3 hours. Taken into basis that the company works 24 hours during 29 days/month, it is calculated that the monthly labor spent for this alternative is 174 hours. Since the unit labor cost is 1.6 $/hour, the labor cost associated to the collection of parts is:

C Labor Cost of Collecting Parts= 1.6 $/hour x 174 hours/month = 278 $/month.

The net present values of the current system and the batch mode are calculated for one year, as seen in Table 1. The costs are assumed to derive at the end of each month and the monthly interest rate is taken as 1.6%. When comparing net present values, it can be seen that the batch mode decreases costs by 46% compared to the current system. This alternative obviously will reduce the quantity of lead alloy required to produce the same amount of grid since the raw material loss in dross will be minimized. Therefore, it will clearly improve the productivity by reducing raw material consumption. This alternative is also advantageous because the waste is reduced at source. As a result, the environmental impact of the company can be considerably reduced by using this alternative.

Separation of Pure Lead from Dross Using Vitaflux™: The dross collected from the melting pot is high in its lead (Pb) content. It is observed that dross formed in the current system contains a quantity of 41% of lead in a purity of 99.92%. As mentioned before, the dross is currently collected and sent to the recycling facility. However, if the pure lead can be extracted from the dross inside the facility, it will be a very beneficiary practice since it will be possible to reuse the pure lead as a raw material. Drossing-off fluxes are used for this purpose in order to accumulate the oxides and allow easy removal from the surface of the molten lead (Brown, 1999, pp. 56-62). Specifically, there is a material called Vitaflux™, which is used to recover the lead from dross at source (NA Graphics, n.d.). Application of the material is very simple. Vitaflux™ is added on the surface of the melting pot before dross is collected. The lead alloy begins to burn with the addition of Vitaflux™ and the pot is mixed until dross becomes fine powder. At the end of the reaction, the dross in fine powder form is accumulated on the surface of the molten metal. Vitaflux™ has been tested in the company and lead was recovered from the dross successfully.

The grid casting department consists of two melting pots. Due to the lead alloy load of the pots, six tubes of Vitaflux™ is needed for each melting pot every time dross is collected. Since dross is collected once a day in the grid casting department, one box containing a dozen of Vitaflux™ tubes is needed every day. In the calculations, it is assumed that one year consists of 350 working days. Therefore, about 350 boxes of Vitaflux™ are needed per year. Results from test showed that the use of Vitaflux™ does not influence the composition of the lead alloy in the melting pot. Besides, when the dross collected after the use of Vitaflux™ is analyzed in the laboratory of the company, it was observed that the pure lead is totally removed from the dross. Therefore, it can be concluded that the use of Vitaflux™ creates significant amount of recovery of lead from the dross.

Experiments conducted in the company for recovering lead from dross showed that the quantity of the dross decreases by 30% when pure lead is separated from dross with the help of Vitaflux™. In other words, this alternative involves 30% of reduction of dross at source. Therefore, even if the impact to the environment can not be quantified, the company's damage to the environment is certainly decreased on a large scale since the dross deriving from grid casting is the highest amount of waste in the company. In addition and more importantly, a significant quantity of pure lead is going to be saved and will be used in production. As a result, this alternative allows a more efficient use of raw materials and a decrease in total recycling costs. The net present values of the current system and the Vitaflux alternative are calculated for one year in Table 2.

In addition to the current system, in the Vitaflux alternative there is a need of extra labor time when mixing the pot and collecting the dross after the addition of Vitaflux. Dross is collected once in 24 hours from 2 melting pots. In the current system, the collection of dross from one pot takes 10 minutes. Taken into basis that the company works 24 hours during 29 days/month, it is calculated that the monthly labor spent for in the current case is 10 hours. Since the unit labor cost is 1.6 $/hour, the labor cost associated to this alternative is: C Labor Cost of Current System= 1.6 $/hour x 10 hours/month = 16 $/month.

The same calculation is made for the Vitaflux alternative taking into basis that the collection of dross from one pot takes 20 minutes. On the other hand, the period in which Vitaflux™ is bought and the quantity of Vitaflux™ needed in this time period is important. It is calculated that 3 month purchasing period gives the minimum net present value. The monthly interest rate is taken as 1.6%. When comparing net present values, it can be seen that the Vitaflux alternative decreases costs by 17% compared to the current system. Since the amount of dross is assumed to be equal to the lead alloy loss, it can be concluded that 3,690 kg/month of lead can be recovered and used in production with the Vitaflux™ application.

Alternative Waste Management Solutions: Setup Wastes Generated From Battery Closing Operation

Minimization of setup time may correspond to reduction of setup wastes for the industries producing sequential products requiring similar production techniques. One way of decreasing setup waste may be to decrease the number of machine adjustments in the productions, where the number of setup waste is constant according to quality regulations. The production scheduling policy used in most of the industries is based on a system where monthly demands are divided into short sub-periods in order to increase the flex-

Table 2. Comparison of cash flow analysis of the current system and the Vitaflux alternative

	Current System	Vitaflux Alternative
Quantity of dross (kg/month)	12,299	8,609
Cost of raw material loss ($/month)	14,605	10,224
Labor cost associated to the collection of dross ($/month)	16.25	32.5
NPV of Vitaflux for 3 months purchasing period	-	19,465*
Net Present Value ($/year)	**158,492**	**130,648**

* Calculated due to data of NA Graphics (n.d., n.p.)

ibility of the production. However this may cause production of abundant numbers of setup waste as a result of inverse proportionality between the length of the sub-period of scheduling plan and the number of machine adjustments. If longer scheduling sub-periods are chosen, then the production of setup wastes and the inventory holding costs will be decreased but the flexibility of the scheduling will be lost. On the other hand, if the scheduling sub-period is shortened, then the scheduling will be more flexible but the number of setup waste will increase. For this reason, there is a need to find out an optimum scheduling plan taking into account the production of setup wastes.

When adjusting the production schedule with the aim of minimizing setup wastes, the primary concern of the car battery manufacturing company is the demand level of products. Products with low demands (products whose monthly demands are approximately equal to the production capacity of one shift) are produced in one or two shifts. Products with high demands (products whose monthly demand highly exceeds the production capacity of one shift) are produced on a weekly basis and are sent to customers each week. Namely, the monthly demand is divided to weekly demands. Since these highly demanded products are not produced continuously, a setup adjustment on the closing machine is needed each time the production of this product starts (each week). The number of setup adjustment will be high, as well as the setup waste generated. An alternative waste management solution is to schedule products so that high-demanded products are produced continuously like the low-demanded products. By this one-month-scheduling method, the monthly demand of high-demanded products will not be divided by four; the monthly demand will be produced in one batch. Namely, the production of these products will continue until the monthly demand is reached.

It is necessary to produce one setup waste while making machine adjustments in the battery closing operation. This setup waste is considered as a refurbished battery and can not be sold to customers. Therefore, the production cost of the battery is considered as the on-site waste management cost of the refurbished battery. One way to establish a pollution prevention approach may be to change the production scheduling policy to decrease the number of refurbished batteries. The relation between the length of the scheduling sub-period and setup waste amount is established to find out the existence of an option to obtain minimum waste. The scheduling sub-period (one week, two week, or four week) in a month affects the number of setup waste generated. Therefore, one of the ways of reducing the setup waste is to determine the sub-period with minimum inventory holding and waste removal cost between different scheduling sub-periods. For this reason, different scheduling sub-periods are analysed, such as one week, two week and four week. The company is currently using a production scheduling based on one-week sub-period. A binary integer programming model is developed to determine the sub-period giving the minimum total cost comprising inventory holding and waste management costs. Inventory carrying within the considered sub-period is ignored. The model is demonstrated below.

$$\sum_{i=1}^{m}\left(W_1R_iX_1+W_2R_iX_2+W_4R_iX_4+\frac{3}{2}h_iX_1D_i+h_iX_2D_i\right)$$

Minimize

subject to

$$X_1+X_2+X_4=1$$

$$X_k\in\{0,1\}\qquad ,\ \mathrm{k}=1,2,4$$

Where, D_i is the demand of product *i*, $W_{i,k}$ is setup waste of product *i* derived with alternative *k*, R_i is the waste management cost of product *i*, and h_i is weekly inventory holding cost of product *I* The model is solved for the given data. Different types of products with varying demands (5 of them are low-demanded and 12 of them are high-demanded) are taken into basis and the forecasts of year 2006 are considered as the monthly demands of each product. The battery production cost associated to the refurbished batteries is considered as the waste management cost of the refurbished battery. The weekly inventory cost of a product is considered as the selling price of the product multiplied by the weekly interest rate which is taken as 0.4%. The number of setup wastes changes according to the sub-period chosen. The one-week sub-period produces 4 setup wastes, the two-week sub-period produces 2 setup wastes and the four-week sub-period produces 1 setup waste per month. The model is solved in Excel Solver for the demands of each month. The option giving the minimum cost is found as the four-week sub-period for each run (In the optimal solution, X_4 is 1 while X_1 and X_2 are 0). The cash flow analysis of the optimal solution of the mathematical model and the current system are compared in Table 3.

Table 3. Comparison of cash flow analysis of the current system and the mathematical model

Costs (YTL/month)	**Current Case**	**Optimal Solution of the Mathematical Model**
January	18.41	0.97
February	18.41	0.97
March	19.30	0.91
April	19.30	0.91
May	15.84	0.89
June	15.56	0.89
July	15.20	0.84
August	14.55	0.84
September	23.15	0.97
October	23.42	0.97
November	24.01	0.97
December	23.42	0.97
Net Present Value ($/year)	**207,240**	**10**

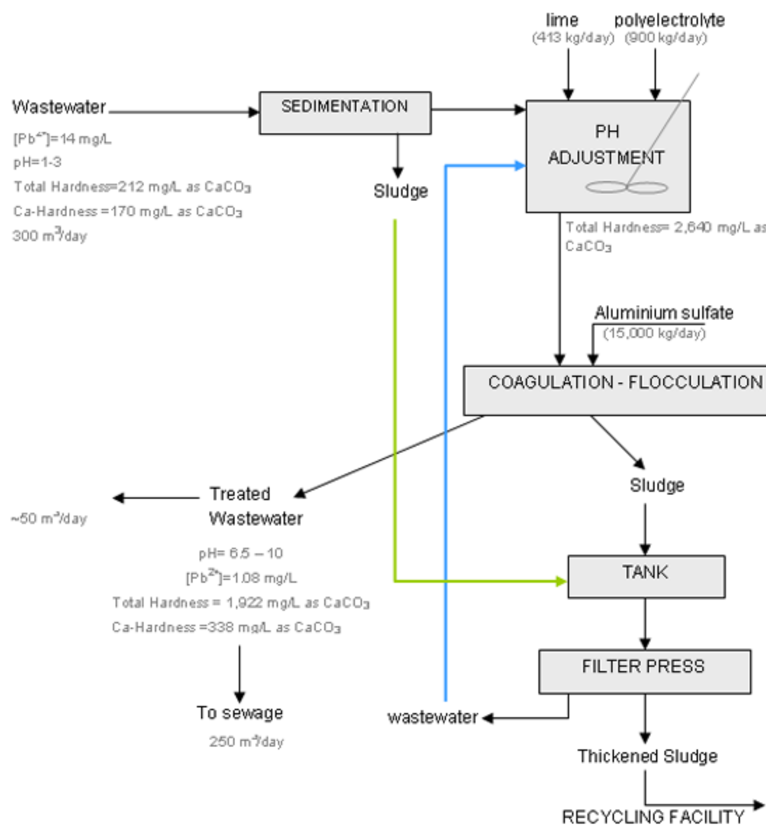
As shown in Table 3, the optimal solution of the mathematical model for each month corresponds to the cost associated to the sub-period giving the optimal solution, the four-week sub-period. When comparing net present values, it can be seen that the four-week sub-period decreases costs by 95% compared to the current system. The wastage of the semi-finished products and the according waste management costs can be significantly decreased with the four-week sub-period. This alternative is an important example showing how the operational decisions can improve environmental problems in a company. However, the production of a high-demanded product in one batch may not be possible in certain cases because of the lead-time of the demands. The operation time of a product may have a long duration, especially when all of the monthly demand is to be produced in one batch. Therefore, the lead time of the other products waiting on the queue will be an important constraint. In addition, production plans must be flexible enough to meet the unexpected demands coming from the competitive environment. The scheduling of this type of environmentally integrated production plan may be difficult and sometimes infeasible for companies that have to deliver products every week.

Modification of the Wastewater Treatment Facility to Reduce the Amounts of Chemicals Used and to Decrease Water Consumption in Production

The cooling and washing water coming from wet-charging process, the charging water of plates and the water used in cleaning equipments are collected and then treated in the wastewater treatment facility operated by the company. Figure 2 shows the unit operations and processes in the wastewater treatment facility of the company.

As shown in Figure 2, the collected wastewater from the resources mentioned above, with a

Figure 2. Operations in the wastewater treatment facility



discharge of 300 m^3/day, is first transferred to sedimentation tank to remove particles, which have high amount of Pb. The sludge formed in this process is sent to the tank used for the collection of sludge. The cleared wastewater (supernatant) is then transferred to pH adjustment unit in which lime (CaO) is added (413 kg/day) to raise pH value of the wastewater from 1.3 to 8.5-9.0. The addition of polyelectrolyte, which is a polymer used to stabilize the sludge, is also accomplished in this unit (900 kg/day). Following up the pH adjustment, aluminum sulfate ($Al_2(SO_4)_3$) is added as precipitant for the removal of Pb dissolved in wastewater in coagulation and flocculation unit. The sludge formed in this reactor is sent to sludge collection tank and then to filter press to become thicker. The wastewater from the filter press unit is recycled back to mixing process. The sludge is sent to the recycling facility since it contains considerable amount of lead. The wastewater from flocculation unit is discharged to sewage while about 1/6th of the volumetric flow is reused as lead suppressant and in negative drying purposes. Since the hardness of the treated wastewater is high enough to stain the containers, it is not possible using it as cooling water or cleaning water for equipments. In cases it is used in wet-charging, white traces will be left on the surface of containers. Hard water is not preferred for cleaning the equipments since it reduces the performance of detergents used. Therefore, using another chemical to adjust pH instead of CaO, for example NaOH, may decrease the water consumption.

The hardness of water is mainly caused by divalent or trivalent ions, Mg^{2+} and Ca^{2+}, and water can be softened by chemical precipitation technique (Patterson, 1985, 220-226). If the chemical (CaO) used for pH adjustment is replaced by another chemical containing an ion, which does not cause any additional hardness, then the treated wastewater can be recycled back to the production system to be used in wet-charging, cleaning, etc. This change is expected to decrease the water consumption of the company significantly. To be able to compare the benefits of using NaOH instead of CaO, there is a need to find out the amount of NaOH required to neutralize the wastewater containing H_2SO_4 and to raise its pH to 8.5.

Calculation of CaO and NaOH Amounts: The theoretical calculation of the amounts of CaO and NaOH to be added to neutralize the wastewater containing 496 mg SO_4^{2-}/L (5.2 x 10^{-3} mol SO_4^{2-}) and to adjust the pH to 8.5 requires the following assumptions:

1. The ionic strength of the wastewater is very low, so that the influence of other ions in concentration estimations can be ignored.
2. The temperature of the wastewater is 25°C.
3. Dissociation of $Ca(OH)_2$, which is the product of a reaction between CaO and H_2O is complete.

$CaO\ (s) + H_2O \rightarrow Ca(OH)_2\ (aq)$ (Reaction 1)

$Ca(OH)_2 \rightarrow Ca^{2+} + 2OH^-$ (Reaction 2)

4. The dissociation of NaOH in H_2O is complete.

$NaOH \rightarrow Na^+ + OH^-$ (Equation 1)

5. The CaO and NaOH used are 100% pure.

Dissociation of H_2O:

$K_w = [H^+]\,[OH^-] = 10^{-14}$

$$\left[OH^-\right] = \frac{10^{-14}}{[H^+]}$$

Mass balance on Ca^{2+} and SO_4^{2-}:

$$C_{T,Ca^{2+}} = [Ca^{2+}] = \frac{\text{number of moles}}{\text{total volume}}$$

$$C_{T,SO_4^{2-}} = [SO_4^{2-}] = \frac{\text{number of moles}}{\text{total volume}}$$

Charge Balance:

$$2[Ca^{2+}] + [H^+] = 2[SO_4^{2-}] + [OH^-]$$

$$2[Ca^{2+}] + [H^+] = 2[SO_4^{2-}] + \frac{10^{-14}}{[H^+]}$$

$$\text{pH} = 8.5 = -\log[H^+]$$

$$[H^+] = 10^{-8.5}\ \text{M}$$

$$[Ca^{2+}] = 5.20 \times 10^{-3}\ \text{M}$$

Therefore, Ca^{2+} to be added:

$$= 5.20 \times 10^{-3}\ \frac{\text{mol}}{L} \times 300\,\frac{\text{m}^3}{\text{day}} \times 1000\,\frac{\text{L}}{\text{m}^3} \times \frac{1\ \text{mol CaO}}{1\ \text{mol Ca}^{2+}} \times \frac{(40+16)\ \text{gr}}{1\ \text{mol CaO}} \times \frac{1\ \text{kg}}{1000\ \text{gr}} = 87.36\ \text{kg}/day$$

The calculated amount of CaO is much lower than the amount added into the pH adjustment tank in the wastewater treatment plant, which is 413 kg/day. The difference is huge, but it could be due to the assumptions made in theoretical calculations especially the grade of the CaO used in the plant, which may be very low. Moreover, the dissociation of CaO and $Ca(OH)_2$ may not be completed in the pH adjustment unit as a result of poor mixing conditions.

Mass Balance on Na^+ and SO_4^{2-}:

$$C_{T,Na^+} = [Na^+] \frac{\text{number of moles}}{\text{total volume}}$$

$$C_{T,SO_4^{2-}} = [SO_4^{2-}] = \frac{\text{number of moles}}{\text{total volume}}$$

Charge Balance:

$$[Na^+]+[H^+]=2[SO_4^{2-}]+[OH^-]$$

The equation was similarly solved and the amount of $[Na^+]$ was found to be 0.01037 M which corresponds to 124.46 kg/day. The reason for adding more NaOH than CaO is the number of hydroxyl ions in the compositions of the chemicals. It should be noted here that there is a significant difference in the solubility of NaOH (1,080 gr/L) and CaO (1.39 gr/L) (Dean, 1992). The NaOH can be dissolved in the water in almost instantly requiring mixing enough to sustain the homogeneity of the wastewater. On the other hand, the solubility of CaO is lower than NaOH.

In the experiments made in the wet-charging operation, it is seen that caustic does not leave white traces in the surface of containers. Therefore it is feasible to use caustic instead of lime. In addition, with the use of lime, it will be possible to treat and reuse the same water for several days. The wet-charging operation uses 150 tones of water every day; so the reuse of the treated water in wet-charging will remarkably reduce the water consumption. The remaining 150 m^3 of refined water may be used in the cleaning of equipments, in the negative drying of plates and as lead suppressant. Table 4 shows the comparison of lime and caustic. Since 100 tones of treated wastewater will be used in equipment cleaning once a day, this amount of wastewater can not be recycled back to production system. Otherwise, there would be a need for tank with capacity of at least 100 tones. Since the water needed in the facility is 300 tones/day, the company will have to use tap water of 100 m^3 to clean the equipments each day. In the calculations, it is assumed that 300 m^3 of tap water is captured in the first day of the week and 100 m^3 of water in each of the remaining 5 working days of the week. Therefore, the weekly water consumption is:

Water consumption = 300 + 5 x 100 = 800 m^3/week.

In addition, as calculated previously, it is also known that 124.46 kg/day of caustic will be used instead of 87.35 kg/day CaO. Since no experiment could have been performed using real wastewater to find out the amount of NaOH for adjusting the pH to about 8.5, the comparison of costs of lime and caustic was made based on theoretical calculation of the amounts of chemicals with the assumptions made. The prices of the pure CaO and NaOH are

Table 4. Comparison of cash flow analysis of the current system and lime alternative

	Current System	**Lime Alternative**
Water Usage (m^3 /month)	7,200	3,200
Total Cost of Water ($/ month)	15,930	7,080
Chemical Usage (tone/month)	2.1 (0.08735 x 6 x 4)	2.9 (0.124460 x 6 x 4)
Unit Cost of Chemical ($/tone)	76.25	562.5
Total Cost of Chemical ($/month)	160	1,680
Net Present Value ($/year)	**174,413**	**94,957**

76.25 $/tone and 562.5 $/tone, respectively. The unit cost of water is 2.21 $/tone. The cash flow analysis of the using caustic instead of lime is compared with the current system in Table 4. As it is seen from Table 4, even though the unit cost of caustic is greatly higher than lime, this alternative gives lower costs because it significantly reduces the water consumption by allowing the reuse of the refined wastewater. A high cost saving is obtained with the use of this alternative. The net present values of the current system and the suggested alternative are calculated for one year, as seen in Table 4. The costs are assumed to derive at the end of each month and the monthly interest rate is taken as 1.6%. When comparing net present values, it can be seen that the suggested alternative decreases costs by 45% compared to the current system.

Comparison of Alternative Waste Management Solutions

Table 5 shows the suggested alternative waste management solutions and the environmental and economical benefits that can be obtained by using them in the company. As seen in Table 5, it was demonstrated that the formation of dross which is the biggest waste problem of the company could be decreased by 30% by using a flux during dross collection. Since the waste was reduced at the source, the waste and recycling costs could be significantly decreased, and most importantly, a very big quantity of lead (Pb) raw material could be recovered consequently. The reuse of the treated wastewater was also a good alternative for the company since a very high quantity of water is used and wasted every day due to production requirements. By using caustic instead of lime, the company could reduce its weekly water usage up to 44%, which would make a considerable decrease in environmental impacts and operating costs.

Table 6 shows the difference in waste management costs when quantity of wastes are decreased according to the suggested alternative waste management solutions. It can be seen in Table 6 that a cost improvement of 63% is achieved by implementing all the alternative waste management solutions suggested. The calculation of the waste management cost of the suggested alternative for the dross obtained in grid casting operation is based on the alternative waste management solution covering the addition of excess parts and rejected grids in batch mode.

FUTURE RESEARCH DIRECTIONS

It is believed that several new approaches and techniques can be added to improve this methodology in the future. A decision support tool can be

Table 5. Suggested solution alternatives and advantages

Wastes	Alternative	Environmental Advantages	Economic Advantages
Dross, excess parts and rejected grids from grid casting	Addition of Excess Parts and Rejected Grids in Batch Mode	Waste reduction at source	Cost Reduction: 94,631 $/year.
		Recovery of 70,848 kg Pb/year.	
	Separate Pure Lead From Dross With Vitaflux	Waste reduction at source. Recovery of raw material from the waste at source.	Cost Reduction: 27,812 $/year.
		Recovery of 44,280 kg Pb/year.	
		Recovery of 1,934 kg Pb/year.	
Refurbished battery from battery closing	Change the Production Schedule	Waste reduction at source.	Cost Reduction: 197,231 $/year
Treated wastewater from the wastewater treatment facility	Use Caustic Instead of Lime	Usage of water decreased by 52,000 m3/year.	Cost Reduction: 79,456 $/year.

Table 6. Comparison of waste management costs

Waste	Current Waste Management Cost ($/year)	Waste Management Cost of Suggested Alternative ($/year)	Cost Improvement
ON-SITE RECYCLED WASTES			
Refurbished Battery	207,240	10	96%
OFF-SITE RECYCLED WASTES			
Dross of grid casting	203,552	108,920	47%
WASTEWATER			
Wastewater	174,413	*94,957	46%
TOTAL COSTS	**585,205**	**108,930**	**82%**

* The addition of excess parts and rejected grids in batch mode is taken into basis

designed for calculating and comparing costs and benefits of solution alternatives. By using such a decision support tool, the unit costs and unit profits can be saved in the system and costs of solution alternatives can be easily calculated. It will be also possible to define criteria and give weights to them in order to compare the operational, technical, economical and environmental characteristics of solution alternatives. Implementation of this environmentally integrated manufacturing system analysis in a different sector or company will also provide important inputs for improving the system and understanding its capabilities.

CONCLUSION

This study has introduced an environmentally integrated manufacturing system analysis methodology for companies willing to be a part of the sustainable development by efficiently using the principles of "cleaner" production techniques in obtaining high productivity levels. The wastes derived from the manufacturing process should be minimized in order to reduce the amount of raw materials used and decrease the need for wastes' recycling. The problem considered is the investigation of a systematic methodology decreasing the waste obtained from the manufacturing process while simultaneously improving the overall performance of the company. This methodology is important in being a roadmap for the evaluation of environmentally friendly and economically favorable alternative waste management solutions. By using this methodology, the manufacturing and waste management system of a company can be analyzed, alternative waste management solutions for decreasing wastes can be investigated, evaluated and compared.

REFERENCES

Akkerman, R., & van Donk, D. R. (2006). Development and application of a decision support tool for the reduction of product losses in the food-processing industry. *Journal of Cleaner Production*, *16*(3), 335–342. doi:10.1016/j.jclepro.2006.07.046

Angell, L. C., & Klassen, R. D. (1999). Integrating environmental issues into the mainstream: An agenda for research in operations management. *Journal of Operations Management*, *17*(5), 575–598. doi:10.1016/S0272-6963(99)00006-6

Boden, D. P., & Loosemore, D. (2007). A technology for production of a cureless paste containing a high concentration of tetrabasic lead sulfate and a low concentration of free lead. *Journal of Power Sources*, *168*(1), 90–94. doi:10.1016/j.jpowsour.2006.10.095

Brown, J. R. (1999). *Foseco non-ferrous Foundryman's handbook, fluxes*. Boston, MA: Buttherworth-Heirmann.

Claver, E. (2007). Environmental management and firm performance: A case study. *Journal of Environmental Management*, *84*(4), 606–619. doi:10.1016/j.jenvman.2006.09.012

Dahab, M. F., Montag, D. L., & Parr, J. M. (1994). Pollution prevention and waste minimization at galvanizing and electroplating facility. *Water Science and Technology*, *30*(5), 243–250.

Dahodwalla, H., & Heart, S. (2000). Cleaner production options for lead-acid battery manufacturing industry. *Journal of Cleaner Production*, *8*(2), 133–142. doi:10.1016/S0959-6526(99)00314-5

Ferreira, A. (2004). Manufacturing improvements in the processing of lead-acid battery plates and reduction in plate cutting with an active-material addictive. *Journal of Power Sources*, *133*(1), 39–46. doi:10.1016/j.jpowsour.2003.12.010

Freeman, H. M. (1988). *Standard handbook of hazardous waste treatment and disposal*. New York, NY: McGraw-Hill.

Gupta, M., & Sharma, K. (1996). Environmental operation management: An opportunity for improvement. *Production and Inventory Management Journal*, *37*(3), 40–46.

Jia, X. (2005). Multi-objective modeling and optimization for cleaner production processes. *Journal of Cleaner Production*, *14*(2), 146–151. doi:10.1016/j.jclepro.2005.01.001

Klassen, R. D. (2000). Exploring the linkage between investment in manufacturing and environmental technologies. *International Journal of Operations & Production Management*, *20*(2), 127–147. doi:10.1108/01443570010304224

Melnyk, S. A. (2003). Assessing the impact of environmental management systems on corporate and environmental performance. *Journal of Operations Management*, *21*(3), 329–351. doi:10.1016/S0272-6963(02)00109-2

Nahmias, S. (2004). *Production and operations analysis*. New York, NY: McGraw-Hill.

Patterson, J. W. (1985). *Industrial wastewater treatment technology*. Stoneham, MD: Butterworth Publishers.

Porter, M. E. (1991). America's greening strategy. *Scientific American*, *254*(4), 168. doi:10.1038/scientificamerican0491-168

Porter, M. E., & van der Linde, C. (1995). Green and competitive: Ending the stalemate. *Harvard Business Review*, *73*(5), 120–137.

Spengler, T. (1998). Development of a multiple criteria based decision support system for environmental assessment of recycling measures in the iron and steel making industry. *Journal of Cleaner Production*, *6*(1), 37–52. doi:10.1016/S0959-6526(97)00048-6

Sroufe, R. (2003). Effects of environmental management systems on environmental management practices and operations. *Production and Operations Management*, *12*(3), 416–431. doi:10.1111/j.1937-5956.2003.tb00212.x

Wu, C. C., & Chang, N. B. (2004). Corporate optimal production planning with varying environmental costs: A grey compromise programming approach. *European Journal of Operational Research*, *155*(1), 68–95. doi:10.1016/S0377-2217(02)00820-2

Xigang, Y., & Zhaoling, Y. (2000). An approach to optimal design of batch processes with waste minimization. *Computers & Chemical Engineering*, *24*(2), 1437–1444. doi:10.1016/S0098-1354(00)00397-5

ADDITIONAL READING

Allen, D. T., & Shonnard, D. R. (2001). *Green engineering: environmentally conscious design of chemical processes*. New York, New Jersey: Prentice Hall.

Bishop, P. L. (2000). *Pollution prevention fundamentals and practices*. Boston, Maryland: McGraw-Hill.

Brennan, L., Gupta, S. M., & Taleb, K. N. (1996). Operations planning issues in an assembly/disassembly environment. *International Journal of Operations & Production Management*, *14*(9), 57–67. doi:10.1108/01443579410066767

Dechant, K., & Altman, B. (1994). Environmental leadership: from compliance to competitive advantage. *The Academy of Management Executive*, *8*(3), 7–27.

Frosh, R. A., & Gallopoulos, N. (1989). Strategies for manufacturing. *Scientific American*, *261*(3), 144–152. doi:10.1038/scientificamerican0989-144

Gianinikos, I. (1998). A multiobjective programming model for locating treatment sites and routing hazardous wastes. *European Journal of Operational Research*, *104*(2), 333–342. doi:10.1016/S0377-2217(97)00188-4

Glavic, P., & Lukman, R. (2007). Review of sustainability terms and their definitions. *Journal of Cleaner Production*, *15*(18), 1875–1885. doi:10.1016/j.jclepro.2006.12.006

Grau, R., Espuna, A., & Puigjaner, L. (1994). Focusing in by-product recovery and waste minimization in batch production scheduling. *Computers & Chemical Engineering*, *18*, 271–275. doi:10.1016/0098-1354(94)80045-6

Guide, V. D. R., Kraus, M. E., & Srivastava, R. (1997). Scheduling policies for remanufacturing. *International Journal of Production Economics*, *48*(2), 187–204. doi:10.1016/S0925-5273(96)00091-6

Handfield, R. B. (1997). Green value chain practices in the furniture industry. *Journal of Operations Management*, *15*(4), 293–315. doi:10.1016/S0272-6963(97)00004-1

Hart, S. L., & Gautam, A. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment*, *5*(1), 30–37. doi:10.1002/(SICI)1099-0836(199603)5:1<30::AID-BSE38>3.0.CO;2-Q

Klassen, R. D., & McLaughlin, C. P. (1993). TQM and environmental excellence in manufacturing. *Industrial Management & Data Systems*, *93*(6), 14–22. doi:10.1108/02635579310040924

Klassen, R. D., & Whybark, D. C. (1999). The impact of environmental technologies on manufacturing performance. *Academy of Management Journal*, *42*(6), 599–615. doi:10.2307/256982

Nevin, R. (2007). Understanding international crime trends: The legacy of preschool lead exposure. *Environmental Research*, *104*(3), 315–336. doi:10.1016/j.envres.2007.02.008

Rosselot, K. S., & Allen, D. T. (2002). *Environmentally conscious design of chemical processes, chapter 10: Flowsheet analysis for pollution prevention in green engineering,* Upper Saddle River, New Jersey, Prentice Hall

Russo, M. V., & Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal*, *40*(3), 534–559. doi:10.2307/257052

Schroeder, R. G. (1993). *Operations management*. New York, New Jersey: McGraw-Hill.

Shrivastava, P. (1995a). Environmental technologies and competitive advantage. *Strategic Management Journal*, *16*(3), 183–200. doi:10.1002/smj.4250160923

Shrivastava, P. (1995b). The role of corporations in achieving ecological sustainability. *Academy of Management Review*, *20*(4), 936–960. doi:10.2307/258961

Spengler, T. (1997). Environmental integrated production and recycling management. *European Journal of Operational Research*, *97*(2), 308–326. doi:10.1016/S0377-2217(96)00200-7

Stefanis, S. K., Livingston, A. G., & Pistikopoulos, E. N. (1997). Environmental impact considerations in the optimal design and scheduling of manufacturing processes. *Computers & Chemical Engineering*, *21*(10), 1073–1094. doi:10.1016/S0098-1354(96)00319-5

KEY TERMS AND DEFINITIONS

Cleaner Production (CP): The continuous application of an integrated preventative environmental strategy to processes, products and services to increase efficiency and reduce risks to humans and the environment.

Eco-Efficiency: A management philosophy that encourages business to search for environmental improvements which yield parallel economic benefits.

Environmental Operations Management (EOM): The integration of environmental management principles with the operations management process for the conversion of resources into usable products.

Pollution Control Methods: Methods covering the elimination of pollution after the waste is generated.

Pollution Prevention Technologies: The modification or redesign of the production process and the introduction of new technologies throughout the product life-cycle, which contributes to the development of new internal routines and know-how's.

Sustainable Development: Meeting the needs of the current generation without compromising the ability of future generations to meet their own needs.

This work was previously published in Green Finance and Sustainability: Environmentally-Aware Business Models and Technologies, edited by Zongwei Luo, pp. 436-461, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 51
Ghabbour Group ERP Deployment:
Learning From Past Technology Failures

M. S. Akabawi
American University in Cairo, Egypt

EXECUTIVE SUMMARY

Ghabbour group "GB Auto," an Egyptian auto trading and manufacturing establishment, has gone through two ERP successive implementations within the past 12 years. The newer implementation has experienced several impediments. The executives and the Board of Directors at the group have thoroughly and aggressively examined the status of the IS services provided by this ERP system and assessed their impact on the quality of decision making at all levels of management. The driver for this was to secure all the necessary platforms and management tools for enabling growth and improving efficiency and effectiveness of the company's business operation and resources. The extent of lack of control and effective utilization of the use of resources in the group has been cited by the top management in many interviews. Following its public offering and registration in the CASE and CMA, GB Auto was legally demanded to provide annual and quarterly audit reports of its varied LOBs' performance. The existing information management infrastructure was not providing such agile services. The trajectory of implementations of integrated Enterprise Information Systems at the group was reviewed in this case study and was duly investigated to assess the effectiveness and appropriateness in servicing those purposes and increasing the company's competitive advantage.

ORGANIZATION BACKGROUND

Ghabbour Group (GB) name has become closely linked to the automotive industry in the Egyptian market – from distribution and manufacturing to after sale services – spanning the full spectrum of the transportation arena. This led the group to adopt the motto "GB Auto - everything on wheels." By the 1990's, privately owned GB group became widely associated with Hyundai passenger cars as a distributor and car assembly manufacturing.

DOI: 10.4018/978-1-4666-1945-6.ch051

The group's role and history in the Egyptian economic field dates back to the 1940's. During this era, the two brothers, Kamal and Sadek Ghabbour established their private auto trading company. The firm was officially incorporated in 1956 as "Ghabbour Brothers." The business area of the firm included trading in automotive products, construction materials, home appliances and electronics.

During the 1960's, the Ghabbours' legacy was firmly maintained despite the economic constraints and hardships in this period, when Egypt embraced the Socialist economic doctrine. By the early 1970's and as a result of the Egyptian government's initiative to adopt the economic open door policy, the group's vision to become serious establishment in the automotive industry not only as a trader, but as a manufacturer, took shape. To that effect, the company started focusing on acquiring licenses and agencies for passenger cars, buses, trucks and automotive parts.

The first manufacturing operation of the group started in 1985 by assembling bus bodies under technical agreement with Scania AB. During this era, Egypt's sole automotive manufacturer was the state-owned El Nasr company for automotive which failed to establish the robust infrastructure for sustainable automotive manufacturing base [removed phrase]. The initial industrial base of GB group started near Cairo-Alexandria Agricultural Road in Qalyubia governorate, which was quite a convenient site due its proximity to major markets and easy access.

With the transfer of management to the younger Ghabbour generations in the early 1990's, the GB group witnessed phenomenal growth and expansion. Horizontal expansion was manifested by the addition of new lines of business (LOB) and diversification of product mixes. While vertical expansion was manifested by increasing the manufacturing base capacities, trade partners and brand names licensing. Hyundai car manufacturer of Korea granted GB group the exclusive agency for the distribution of its passenger cars in Egypt since 1992.

In 1995, GB group was also granted by Hyundai the technical license to assemble passenger cars - Completely Knocked-Down (CKD) units in addition to trading in the Completely Built Units (CBU). Since then, GB group acquired additional agencies for Mazda passenger cars, two and three wheelers products of Bajaj of India; assembly licenses for Mitsubishi trucks and buses; and Volvo trucks and buses. In 2006, GB group entered the transportation services area, for both passengers and cargo, by establishing the new subsidiary - Haram Company.

In 2003, Chairman of the GB Group, Dr. Raouf Ghabbour, established the Raouf Ghabbour Investment Company (RGI) as the holding company of the group's subsidiaries operating in varied LOB portfolio. A list of these subsidiaries is shown in Table 1 together with the business activities pertaining to each subsidiary. In Table 2, we list the various corporate business organizational units in the group and the associated business activities each performs. Those corporate business units provide services across the entire functional activities of the group's subsidiaries.

Before year 2006, RGI (S.A.E) operated as the holding company for GB group's affiliated companies which were wholly owned by RGI. However, one of the groups' affiliated companies – GB Capital was subsequently named as the holding company for the group and RGI became in effect a wholly owned subsidiary. In April 2007, GB capital name was changed to GB Auto.

In July 2007, GB Auto made its Initial Public Offering (IPO) and applied to the Cairo and Alexandria Stock Exchange (CASE) and the Capital Market Authority (CMA) to become a share-holder company with the name of GB Auto (S.A.E). The Ghabbour family maintained ownership of majority of the shares (71 percent). However the offering attracted many foreign and local share holders. A board of directors was formed soon

Table 1. GB holding company and subsidiaries

Ser.	Company Name	Business Activities	Number of Employee
1	RG Investment S.A.E (RGI)	Holding Company for Consolidated Subsidiaries (prior 2006)	101
2	International Trade Agencies Marketing Corp. S.A.E (ITAMCO)	Sales & Dist. Of CBU & CKD PC unites. Spare Parts Dist. Operation of PC after Sales Service Network.	1444
3	Interland Motors S.A.E.	Sales & Dist. Of Volvo heavy range trucks.	33
4	Egyptian Vehicles Manufacturing Company (Ghabbour Misr) S.A.E	Assembly of mini-bus and large coaches, operation of Commercial Vehicles after sales service network and spare parts. Distribution of Volvo Construction equipment	1779
5	Haram Transport Company S.A.E.	Operation of passenger and Cargo Transportation Service	904
6	Cairo Individual Transport Industry CITI S.A.E	Operation of motorcycle and scooter (2 & 3 wheelers) LOB.	64
7	Prima Engineering Industries S.A.E. (PRIMA)	Assembly of Hyundai CKD PC and Mitsubishi light Commercial vehicles, Manufacturing Semi-Trucks	45
8	Vehicles Components Industries S.A.E (VCI)	Manufacturing Semi-Trucks and super-structure, as well as PC Components	45
9	Ghabbour Continental Trading Company free Zones (Alex) S.A.E	Distribution of PCs, mini-buses, spare parts and tires from Alex FZ	17
10	Engineering Company for Marketing and Trading S.A.E	Distribution of tires	33
11	GB Buses, S.A.E (GE)	Manufacturing of bus and Coach bodies	134
12	GB Auto	Holding Company after 2006	1

Source: http://www.ghabbourauto.com

after, and Dr. Ghabbour assumed the position of the Chief Executive Officer of the group.

In September 2007, the group applied to the CMA and the Ministry of Investment for possible merge of its eleven subsidiaries into two major subsidiaries keeping the smaller free zone companies out of this merge. This merge was legally effective by January 2008.

Business Growth Outlook for the GB Group during the Past 5 Years

The historical business growth trajectory and performance of the GB group before the public offering is one of the few business success stories by the Egyptian standards. The group's workforce has also witness substantial growth during the

Table 2. Corporate cross-functional business unites in the group

Unit	Activities
Finance & Accounting	Manages & administer all the group's subsidiaries and the holding company's accounting and finance
Supply Chain	Control all the group warehouses, purchasing and logistics
Internal Auditing	Performs the audit function for all the group's activities in all subsidiaries
Legal Affaires Information Technology	Administers all legal affairs for the group Provides and Manages IT/IS Services to the groups' companies
Administrative Services & HR	Human Resources administration for the group

Table 3. GB auto results for the full-year 2008

(LE million)	FY2008	FY2007
Passenger Cars Revenues	3,675.5	3,314.4
Commercial Vehicles Revenues	740.9	590.0
Motorcycles & Three-Wheelers	571.3	528.2
Other Revenues	204.7	197.5
Total Sales Revenue	5,192.4	4,630.1
Gross Profit	872.3	670.2
Gross Profit Margin	16.8	14.5
Selling & Administration	-277	-218.8
Others – Income / (Expenses)	32.2	18.3
Operating Profit	627.5	469.7
Net Provisions	19.0	112.4
EBIT	646.5	582.1
Foreign Exchange Gains (Losses)	-18.3	2.8
Net Finance Cost	-116.2	-98.4
Earnings Before Tax	512.0	486.5
Taxes	-94.1	-50.7
Net Profit Before Minority Interest	417.9	435.8
Minority Interest	-2.0	-2.3
Net Income	415.9	433.5
Net Profit Margin	8.0	9.4

last four years. This growth is manifested by the nearly tripled human resources tally employed by the group in the period from 2004 to 2008. Table 3 exhibits GB Auto Corporate financial results for the year 2008, while Table 4 shows the increase trend of the employed workforce according to the various broad areas of business activities.

The revenues breakdown of the group from the various broad lines of business as identified above is shown in Figure 1 for FY07.

The last graph identified several lines of business which comprised the following products.

1. Passenger Cars (Hyundai-Built CKD and CBU).
2. Commercial vehicles including (Mitsubishi buses and trucks, Hyundai buses, Volvo trucks and other super structures).
3. Two and three-wheelers of Bajaj brand/ models.
4. Others, including material handling equipments for Linde brand of Germany, tires for Lassa brand of Turkey and Double Coin brand of China, Volvo brand construction equipment as well as Monroe brand shock absorbers.

Management Structure of the Case Firm

GB group's management structure has seen constant changes in response to the changes that occurred during its trajectory of business evolution. We will not indulge in presenting these evolutionary developments. Rather will highlight the major organizational and structural changes before and after the IPO milestone which represented major phase change in the GB group's business practices.

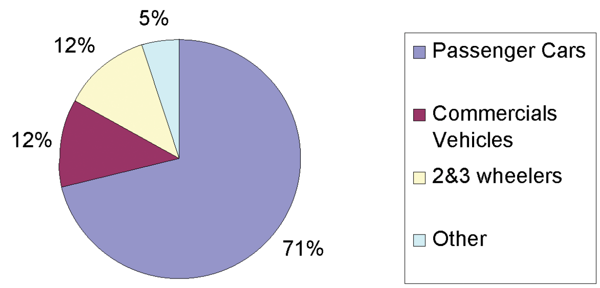
Prior the public offering, GB group has been managed through a hierarchical structure typical

Table 4. GB workforce growth within 2004 - 2007

Business Activity	2004	2005	2006	2007
Passenger Cars	1021	1702	1998	2321
Commercial Cars	624	887	1308	1518
Scooter & Motorcycle	22	33	57	61
Others	92	193	806	1096
Central / Corporate	615	697	1041	1144
Total	2392	3512	5210	6140
Annual % increase	-	46.82%	48.83%	19.3%

Source: http://www.ghabbourauto.com

Figure 1. Revenues breakdown by LOB (Source: http://www.ghabbourauto.com)



of a privately owned company. Dr. Ghabbour, the chairman, was the ultimate decision maker in any major group–wide decisions. The group was run by a number of senior administrators – reporting directly to the chairman. Senior administrators' organizations were identified according to the particular LOB in which their organizations operate. For example, passenger cars sales organization was led by a vice-president who has control over all the related activities of passenger cars, such as sales, marketing, after sale services and warehouses in addition to the company's showrooms.

Likewise, commercial vehicles which by nature spanned several products, was divided among several organizations; each concerned with a particular product line, for example bus sales, marketing and after sale services were headed by a sector director. While heavy trucks sector was assumed as a different organizational unit with a different sector head.

The functional sector divisions are provided with corporate services that cut across the whole group's functional units and subsidiaries. These services are provided by the following divisions:

- Finance and Accounting
- Information Technology
- Legal Affairs
- Logistics and Supply chain
- Human Resources and administration
- Internal Audit

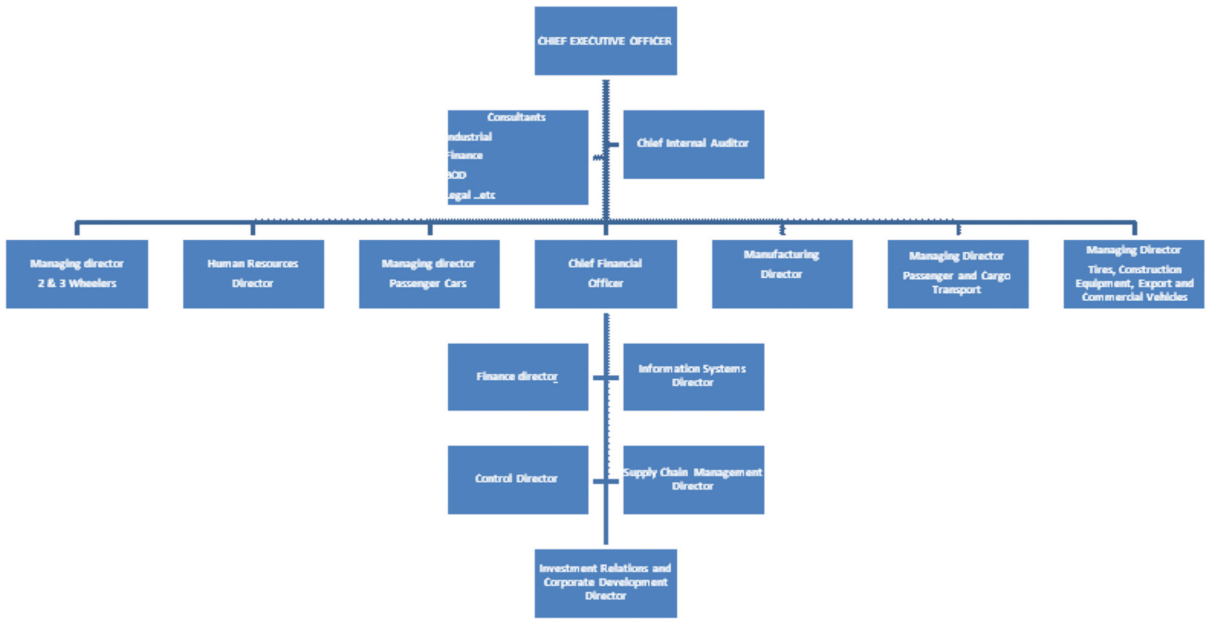
In addition, industrial and construction projects were treated as corporate projects under the direct administration of the group's chairman. Upon the incorporation of the group subsidiaries and the holding company as a joint stock company with the name GB Auto S.A.E, the group has witnessed major organizational restructuring. This resulted in promotions of division heads and the recruitment of key corporate positions such as a Chief Financial Officer (CFO) and investors relations corporate development director. Further Dr. Raouf Ghabbour was named the CEO of the incorporated company. A board of directors (BOD) was formed composed of share holders representatives and industry and finance experts. The GB Auto's organization chart by mid 2007 is exhibited in Figure 2. The management of such a wide range of activities performed by the group cannot be easily described according to the classical organizational theory. However one can postulate that the group is organized according to business sectors as well as product divisions with centralized support service sectors providing cross-functional services to other business sectors and product divisions.

SETTING THE STAGE

Research Method

In this case study, we examine the use of enterprise resources planning (ERP) system in a longitudi-

Figure 2. GB auto organization chart in 2007



nal case study to assess the factors impacting the success and failure of adopting those systems within the context of a developing country's business, legal and political environment. ERP systems are often regarded as an organizational "*administrative framework for planning, conducting and monitoring a large array of functionally segmented operations in a way that both accommodate, in real time, the intrinsic cross-functional interdependencies underlying these operations.*" This postulate implies that '- [ERP systems] *are packages particularly germane in influencing human agency at work*", Kallinikos, (2004, p9). The use of those systems will require making changes in the workflow of the business processes, either through redesigning the old ones to fit the ERP's embedded "best practice" or the re-alignment of the existing workflow. Moreover, ERP systems cut across all levels of management. They are used as the transaction processing workhorse, populating the ERP database with transactions initiated from various business units. The intent is to have one unified enterprise view of the organization's business operations. The prerequisite of such condition is to ensure the timeliness and integrity of the resulting database. To that end synergy among the operational units in organizations has to be provisioned to all stakeholders through a real-time procedural platform. Some members of the user community act as creators/initiators of data items, others as consumers, with each alternating role according to the business model ensued. Such a network of use role-playing will put the ERP operation at the center of the organization's information management model.

The information systems (IS) research on the use of ERP systems in organizations followed two approaches: variance- and process-based approaches (Mohr, 1982). The variance approach included two streams. The first focused on identifying the critical success factors (CSF) that facilitate implementation and "going live" of ERP systems (Bancroft et al, 1996; O'Leary, 2002; Ptak, et al, 2000). The other focuses on development of constructs that assess impacts of ERP systems on work practice (DeLone, et al, 2003; Seddon, 1997; Sedera, et al, 2004; Ifinedo, et al, 2007; Ross, et al, 1999). Conversely, a process based approach focuses on understanding the dynamics of change "how change emerges, develops, and diminishes over time" (Markus, et al, 1988).

We examine the dynamics and success of adoption of the ERP's "administrative framework" during the deployment of two successive implementations at GB Auto, the firm studied in this case. A longitudinal multi-year research project is thus conducted using variance research approach. The study was conducted in two phases, the first, taking almost three years, included the analysis of the dynamics of an operational ERP system, in production for nearly nine years. Phase two of the research began after the "go live" of the second implementation and took one year.

Longitudinal case study approaches are well established in social research. Qualitative panel or cohort studies are commonly used when measuring change over time, whether in attitudes, behaviors or experiences and when looking at causal links. Moreover, the dynamism of the techno-social world points to the need for dynamic methods of enquiry. People's perspectives on technology artifacts use are not fixed and are liable to change for a multitude of reasons. Consequently, research methods with potential to capture this fluidity may be more illuminating than other approaches. These circumstances create a unique role for longitudinal qualitative research which can provide rich information on people's perspectives and how and why these are perceived to have changed over time. The contemporary longitudinal study involves repeat follow-ups of a single sample, panel or cohort and is the approach followed in this research.

To that end, focus group forums from GB Auto, the case firm, were formed to gain understanding of the implemented ERP's operational dynamics during the post implementation period. This is not to be confused with "action research" methodology since no attempt was made by the researcher to make changes in the operation during the study. Those forums enabled direct observations of the working of the system and conducting interviews with all stakeholders. Thirty focus group sessions for different business units were conducted. Each session lasted three to four hours with average participation of 10 to15 persons comprising representatives from business unit management, system users and IT specialists. The researcher acted as session mediator with selected business area specialists as subject-matter process references. Session note-taking and minutes were administered by an MBA student who worked as an intern with the firm.

Theoretical Perspective

Many researchers examined the impacts of the success/failure factors on the accrued value of the ERP systems use to the adopting firms (e.g. Sedera et la (2002), Gable et al (2003), Sedera et al (2004), Ifinedo (2006a), (2007)), and the extent of success in adopting those systems at multiple organizational levels. In this research we pose the following question: "*how relevant those success factors are for the adoption of ERP services in the evolving business organizations within developing countries (DC) business culture"?* The work of Sedera and Ifinedo stands particularly relevant to this scope. To that end, this research is divergent from their works on two aspects: a) cultural context, and b) methodology and analysis tool. On the one hand, the work of Sedera (2002) focused on assessing ERP post-implementation success factors across the organizations' stakeholders (employment cohorts) within Australian Public sector Organizations, and used "*a dual survey approach*". The statistical sample used comprised of 27 Queensland Government agencies, running live SAP systems. While Ifinedo and Nahar (2007), on the other hand, addressed, more or less, the same research domain within the Baltic-Nordic region in Europe. Their sample survey focused on private sector organizations, again using survey measurement instrument which targeted around 500 firms in the Baltic-Nordic region. Conversely, this research is conducted within the

Egyptian business domain and targeting the same research questions addressed previously, using longitudinal Case study approach, as opposed to survey methodology. Qualitative analysis was used for the study of the complex dynamics of the techno-social interaction during the ERP post-implementation phase. The use of longitudinal Case approach was chosen due to several reasons. First, comprehensive and precise information on the rate of penetration of the ERP use in the Egyptian business market proved hard to compile. Second, ERP vendors' statistics on their respective ERP product deployment in the Egyptian firms were extremely misleading. During our study, we noticed there was no differentiation between use of limited number of modules and the use of the full ERP suite modules. Such shortage of information defies the objectives of analyzing success factors for the category of Enterprise-wide IS. Third, the extremely low survey responses, the method that is commonly used for IS variance studies in more advanced and developed countries, discouraged conducting survey data collection within the less matured DCs environment. Instead, as contended earlier, the focus group approach in a longitudinal case setting was used involving all stakeholders in the Case study.

Review of Literature

The two successive ERP implementations in this case falls in the class of commercial off-the shelve software. Those configurable packages are perceived as embedding "best practice" of business processes and the ensued workflows. Often, the objectives of companies in adopting this strategy in managing their information processing needs are to seek cost reduction, increasing productivity, improve customer satisfaction and relations with their suppliers, and ultimately improve their competitiveness in the global networked market space (Davenport, 1988). Due to the far reaching impact of such implementations, many researchers have focused on the identification of the critical success factors (CSF) which facilitate successful implementation and "go live" of the system.

Further, successful adoption of ERP systems is often viewed from the perspective of IS success measurement model developed by Delone and McLean (1992, 2003). This model postulated six independent dimensions. Each dimension provided a measurement construct for the assessment of the extent of success of the IS use. Those six constructs were viewed from the perspective of IS use at large. Many researchers argued that, due to the specific nature of the ERP systems, D&M's model would have to be modified to encompass these peculiarities (Seddon, (1997), Sedera et la (2002), Gable et al (2003), Sedera et al (2004), Ifinedo (2006a), (2007)). For example, Gable, et al (2002), projecting those measurement constructs on ERP, argued for limiting the dimensions to four: a) Information Quality, b) System Quality, c) Individual Impact, and d) Organizational Impact; eliminating the User Satisfaction and Use dimensions in the original D&M model. Ifinedo (2007) then extended those four dimensions by adding the two constructs: Vendor/consultant Quality and Workgroup Impact. Fryling, (2005), in attempting to accommodate the special character of ERP project implementation, advocated "*user acceptance and IS success are highly influenced by the user community for which the ERP system is intended. The earlier a user is involved in the process the more likely they will ultimately be satisfied with the ERP and the more likely they will actually use the system*". This assertion considered the original D&M IS success model as relevant only during the implementation project execution. The novelty of the extensions made by those researchers for the ERP systems success lies in the recognition that ERP implementation and deployment is multi faceted encompassing technological, operational, managerial, strategic, and organizational related components. In particular, the inclusion of the workgroup factor by Ifinedo (2007) as an additional dimension is striking, since successful implementations in large firms depend largely on availing reliable communication infrastructure to empower collaborative work among the user community.

Other critical factors that were researched also in relation to ERP implementation and deployment are the information quality dimension and system configuration. For example, Xu et al (2002) and Madapusi et al (2007) alluded to factors such as top management support, training, communication, employee relations, project team, quality controls, and change management, being crucial in achieving high ERP information quality. The impact of ERP system configuration and/or setup revisions and enhancements of the ERP on the deployment success were also studied by Fryling (2005), Nicolaou (2007), Light (2001), Mensching (2004) and Nah (2001a). Those studies asserted that both the nature and timing of system changes are significant factors to be considered in this respect. As noted in the case, those factors have not received much attention. However, most ERP practitioners considered these factors having profound impacts in the post implementation phase. They assert the importance of the Vendor/Consultant quality factor postulated by Ifinedo (2007). As Fryling (2005) asserts and from our experience, ERP products are continuously subject to software maintenance in the form of patch fixes, version upgrades, add-ons and licensing policies changes by the ERP vendors. The likely disruptions in the operations in such mission-critical applications due to such incidents will undoubtedly send mixed signals to all stakeholders involved. Management will be forced to consider what measures should be taken to secure business continuity, for example through investing in operational redundancy to achieve 24x7x365 operation mode. End-users will have to be alerted and retrained because of the implications of the fixes and/or upgrades on the look and operational changes upon applying these upgrades. IT organizations will alternate between working in project operation mode that these upgrades require, constrained by money, people and time; and at the same time guarantee the "business as usual" systemic use mode of the ERP system by the organizations.

Moreover, it is observed that implementation and utilization of ERP systems in organizations represent a radical change from the operation of legacy systems as noted by Brazel and Dang (2008). ERP workflow crosses functional boundaries that exist in large organizations. The changes Ifinedo (2007) proposed to include the workgroup parameter to D and MIS success model are recalled here to manifest this. In the context of ERP adoptions, operational (rather than technical) factors were also addressed by many researchers. For example, factors such as: business operations coverage of the package and the number of licensed users (Francalanci, 2001; Kumar, et al., 2001; Markus, et al., 2000b; Parr and Schanks, 2000), the implementation process details (Sawyer and Southwick, 2002). The impacts of organizational and national cultures (Krumbholtz et al., 2000; Soh, et al. 2000), users and organizational learning (Fleck, 1994; Parr and Schanks, 2000), knowledge transfer from suppliers to the host organization (Lee and Lee, 2000), the configuration and setup of the package's parameter tables and their overall impact on the architecture and flexibility of the ERP packages for rapid adaptation (Fan et al., 2000; Spott, 2000). Some articles also addressed the impacts of ERP process standardization and the restructuring of organization tasks or business processes reengineering (Kumar and Van Hillgersberg, 2000). The entanglement of the ERP technology with other technological artifacts, such as networks, servers, workstations, etc., as well as commercial interests and external social practices have been addressed by Ciborra et al., (2000), Ciborra and Hanseth, (1998), Fleck, (1994), and how the interdependencies of those components impact the implementation and adaptation in organizations.

CASE DESCRIPTION

Information Systems Services and IT Deployment in GB Group

The deployment of IS/IT services in GB group has seen evolutionary developments since the early 1990's. This evolutionary process can, roughly speaking, be divided into three overlapping phases:

Phase 1: Early adoption phase, which spans the period from year 1990 to1997.
Phase 2: First ERP implementation period, covers the years 1997-2006.
Phase 3: Second ERP deployment phase, which covers the period from 2004 to today.

It is worth noting here that documented records on the details and nature of the IS/IT utilization in phase 1 are missing. This is due to the disappearance of the technical staff for the support of these services and/or the non-existence of physical archives that one can reference to enable the analysis of the usage nature during this phase. In order to shed light on the nature of the use of IT/IS in the group during this era, we have turned our attention to collect data from IT service providers of the group. Also, we asked old timer employees in the finance, accounting and IT divisions to help in identifying the nature of use of IS/IT in the group during this period.

Based on those sources and the remnants of the old systems that were found in the various affiliated companies, the next section provides the foundation of the IT/IS service map in the group during phase 1.

Phase 1: Early IT/IS Deployment

We can assert here that the first serious IT/IS services deployment in Ghabbour group dates back to 1994. Two of its manufacturing subsidiaries, namely PRIMA and VCI companies contracted ORASCOM firm, an Egyptian technology firm, for the deployment of the MFG PRO package of USA-based QUAD Company. MFG PRO runs the following application modules in both factories:

- Inventory management;
- Purchasing;
- Manufacturing, which included
  - Bill of Materials (BOM)
  - Routing
  - WIP
  - Quality Control
- Financials, which included
  - General Ledger
  - Accounts Payable
  - Accounts Receivables
  - Fixed Assets
- Invoicing.

The package was in operation for the period from 1994 to 2003, running on HP 9000 server using HPUX, the HP UNIX variant operating system platform. In an interview with Engineer Albair Shafik, IT specialist in the Ghabbour group IT department, who prior working for the group, was the resident consultant of ORASCOM responsible for the implementation of the MFG PRO package. The following account on the implementation process of the package in the two targeted sites is given next:

"Soon after the contract between ORASCOM and GB group was signed, I was assigned the task of the lead implementation consultant in 1994. We started the first implementation with PRIMA site which lasted for 12 months. We followed this with the VCI site, which lasted for another 6 months. The two factories were new and there were no old IT systems to match. The team started a kind of "clean slate", applying the standard MFG PRO processes and procedures. This made our task much easier and controllable. We have to upload all the master data for the items, suppliers and customers from manual documents. We developed coding system for the master databases, which

was arbitrarily coined without regard to any particular standard. There was even no coordination between the coding systems adopted by the other subsidiaries. The main feature in those two implementations is that we did not do any major customizations and when we needed such customizations, the task was wholly outsourced to the implementer, ORASCOM. Since the implementation and operation of those two factories, they operated satisfactorily until they were migrated to the BaaN system in December, 2003. In my view, this implementation was very successful because it fit properly the user requirements on the one hand. The user did not contest the new system operation and adopted the package's inherent "best practice" processes without much resistance".

In his opinion, this system survived for nearly nine years and gained acceptance and stability due to several factors, among these are:

- Extensive user training was carried out in the early stages of the implementation;
- GB group's MFG PRO implementation was among the first "customers" using the package in Egypt. This has prompted ORASCOM to fully support it in order to make a success story and identify it as its flag-ship in the course of marketing this package;
- The implementer has gone through a thorough and detailed business requirements analysis which helped in properly configuring the package to suit these requirements;
- The system support was totally outsourced after the "go live" for some period, and with the eagerness to make a success story, the vendor exerted every effort to operate the system efficiently and effectively.

Concurrent with the implementation of MFG PRO, ITAMCO and Ghabbour Egypt subsidiaries implemented a sales and distribution package called LEADER, which run on Windows server platform. A third package called El Mokhtar was also implemented for the notes receivables (NR) application for the finance and accounting department. These two implementations were run and operated by separate IT units in each beneficiary division.

This phase of IS/IT deployment can be truly characterized as initiation stage and the implementation of technologies can only be described as fragmented and as such, has caused a great deal of headache for the company when senior management decided at last to streamline all systems under one company-wide unit.

It is an interesting observation to note that even though there was a great deal of common data between the different subsidiaries, each division created their own master databases and respective coding system. This, of course, resulted in duplication of databases among the subsidiaries, thus creating inefficiency within the divisions and increased difficulties when the decision was made to integrate all systems. Also, a great deal of time and money was wasted on the company's behalf. If all relevant information had been shared from the beginning, it would have saved the GB Group a large sum of money and a great deal of time.

The separation of the master databases (customers, suppliers and items) led to the adoption of non-standard and non-coordinated code. The non-standardization of both data structures and processes was deeply rooted in the group's culture. This impeded the subsequent enterprise-wide IS/IT adoption and created mounting problems to the IT organization in trying to fulfill the integration among the group's subsidiaries.

The fledgling islands of computing in the group's companies have taken the IS/IT adoption curve to the second stage of Nolan's (1979) stage model, deepening the non-standardization ailment and strengthening the "silos" mentality in the group's IT management. This particular issue had the gravest impact on the implementation of the first ERP system across the entire group. It dictated an architectural design of this implementation that led to its failure and eventual abandonment as will be seen later.

The senior management of the group realized the imminent increased dependency on the use of IS/IT services all over the group's companies. A decision to centralize those services under one single group-wide unit was thus made. A large enough system to service not only the existing companies but any further expansions in the group's lines of business was to be acquired.

Phase 2: From Expansion to Control, the First ERP Implementation Era

By the year 1997, GB group top management invited prospective systems providers to bid for the procurement, installation and implementation of integrated information systems to serve the needs for the entire group's affiliated companies. Several proposals were presented. Afterwards, the group accepted an offer submitted by a Saudi company named Integrated Systems International (ISI). ISI, the sole agent for BaaN ERP in the Middle East, was chosen to provide the full spectrum of the BaaN application modules. These modules included: manufacturing for discrete and process industries, distribution, finance, service and maintenance, project, transportation, Business Organization Dynamics, Enterprise Modelling and Orgware.

A central IT unit was established as a corporate unit serving the entire group's affiliates. The unit was mandated to install the necessary infrastructure needed for the implementation and operation of the BaaN ERP system. The IT division subsequently contracted IBM branch in Egypt for the acquisition of the hardware, system software and database management system in preparation for the BaaN implementation. The initial configuration installed as the platform for the ERP system operation consisted of the following:

- Three RS/6000 IBM servers, one model H70 to be located in Kalyoubia site and two machines, H70 and H50 models to be located in ITAMCO company at Abo-Rawash site;
- AIX OS, IBM's Unix operating system variant was used for all servers;
- Initially, the T-Base database management system (DBMS) engine was installed, which was replaced in 2001 by the DB2 DBMS

In 1998, BaaN release IV was successfully installed. A fascinating point of interest in this case is that even though BaaN release IV provided an option for group configuration, GB Auto decided to forgo that opportunity in favour of speeding up the implementation process.

The implementation project was executed in 3 stages; the first started with the trading companies and took three years to complete. Thus, in 2001, implementation had taken place in the following companies: RGI, Pre-custom Company, GCT (free zone), GIT (free zone), EMT, ITAMCO, and INTERLAND. Stage two took place during the period of 2001-2003, and resulted in five companies' implementation of BaaN: Ghabbour Egypt, GE Manufacturing, PRIMA, PRIMA Inventory, and VCI.

What is important to consider in those implementations is that from the start, integration was never deliberated. Each company was to be configured as a separate instance with all its master data- even though there was great amount of common data (customers, suppliers and items) among those instances. For example, the customer master database was separately compiled for each individual company regardless the repetition. Another issue that is evident in this case is that the analysis of requirements done prior to the implementation was never thoroughly documented. Also, there was a great deal of customizations taking place in the implementation process in order to fit in the non-standard business processes of each individual company. These customizations have provided nothing more than problems in the long run. In essence, they only encouraged non-integration between the subsidiaries and led to multiple fundamental operational difficulties. The vendor's support was greatly lacking and the group's relationship with them became strained. Due to these problems, as well as others, the IT division had to rely solely on in-house experts who were unable to keep up with the frequent errors due to their limited knowledge on this system.

These problems, as well as the difficulties inherent in the system usage practices led to a fundamental distrust in the system. Data manipulation became commonplace, bringing about the issue of data integrity and validity. This is how the "Excel" culture of Ghabbour Auto came about. As the program aged, costs rose and it became evident that the BaaN operation had come to a dismal conclusion. As Engineer Adel Nassem, the then newly appointed IT and Business Director in 2002 recalls:

"In 2002, when I took charge, the level of business relationship with the BaaN system vendor representative in Dubai was so strained that I had to call BaaN headquarters in Europe asking for help and their interference".

Help was indeed received, BaaN HQ diverted Mr. Nassem's support request to India subsidiary. As a result, GB group contracted two Indian specialists to be on board for fixing the problems of BaaN system which was falling in a vortex spin.

The situation gone from bad to worse by deepening the extent of customizations, forging the non-integration and prohibited scalability and/or upgradeability of the system. Basic functionalities such as consolidating financial reporting were never achieved. Period closing (quarters and/or yearly) was never achieved and data fixing was common practice without regard to audit rules and/or accounting best-practice. In short, what the group ended up with by 2004, a disparate database repository that is totally open for manipulation with no audit trails and no confidence in the validity of the data values.

By 2003, stage three of the implementation project started and the MFG PRO package's operation at both PRIMA and VCI manufacturing installations were stopped. Their databases were migrated to BaaN system, still as separate companies. What is intriguing is that no attempt was made, during this migration process, to perform any form of data cleansing, standardization of the coding systems and/or the processes. This has furthered the "silo" IT operation feature, albeit using the same BaaN "ERP" infrastructure.

As a result of this data chaos, the professional staff in the various functional areas (finance, accounting, logistics, manufacturing, etc) developed the "evil" habit of extracting the data from the BaaN database to their Excel sheets. They performed whatever manipulation they need on this data without regard to audit rules and/or validation and verification controls. To respond to the management demands for performance reporting and to enable fast reply to ad hoc queries whenever decisions are imminent, the "Excel" culture proliferated. Data from BaaN was continuously uploaded to individuals' PCs. This led to creating multiple pictures for the same fact violating issues of data integrity, security and increasing the risk of loss of resources. As engineer Nassem contends:

"The BaaN system has become totally open to the extent that any one from any department could make changes that may be reverted next day just to cover up some mistakes or fix some figures"

The operation of the BaaN, besides its functional flaws, has come to a disastrous and risky juncture when the cost of operating the hardware platform reached prohibitive limits. At the start of the BaaN operation, in order to secure safe and reliable operation of the system, service contracts were signed with IBM for the maintenance of the hardware, systems software and the DB2 DBMS. With the aging of the hardware, the support contract bill rose to extremely high values with no evident ROI. The IT division decided to sever this contract and opted for another hardware support contract other than IBM with the implications of lower performance. The same policy was also extended to the BaaN support contract with the risk of not having support at all for this package, this led to the division becoming under siege and dependent on the foreign (Indian) hired specialists

To worsen the situation even further, with no in-house know-how of database house-keeping, the database sizes for some instances (companies) reached critical levels (85percent to 90 percent limits) requiring major re-configuration process. A situation when referred to the BaaN vendor was unprofessionally exploited by the vendor to recover lost revenues.

Technology Concerns

At this point in time, Dr. Aram Youssef was recruited as the new Vice-Chairman of the group. Dr. Aram felt that it was time to seek the services of external consultants in order to establish where the group went wrong. As a result, a team from the American University in Cairo was consulted. The consultants were requested to evaluate the current IT/IS infrastructure effectiveness and the services provided by the applications. And to propose solutions and recommendations to help the group establish a responsive information systems services. The team met with the group's Chairman, Dr. Raouf Ghabbour; the Vice-chairman for support functions, Dr. Aram Youssef and the IT and Business development Director, Mr. Adel Nassem in February 2004 and were requested to investigate the current IT/IS infrastructure and applications problems and propose recommendations and solutions for the group to establish a robust and responsive information systems services. The team spent eight weeks in activities related to fact finding and data gathering, analysis of the existing IS utilization and other issues related to the operation of the BaaN system effectiveness. The team also surveyed the supporting IT infrastructure and the IT organization unit. At the end of this period, a report was presented to the group's senior management with a proposed action plan to establish a robust and responsive information systems infrastructure. The proposal called for the rejuvenation of the three building blocks of IS building blocks in the group: the systems platform (hardware and system software), the network and communications infrastructure and the business application software platform. The chairman of the group, Dr Raouf Ghabbour, approved the recommendations and the plan was carried out accordingly as will be discussed later.

The consulting team's diagnostic report made several observations explaining the reasons and factors that made the process of BaaN implementation reach a dead end and finally abandonment. These observations were in complete congruence with most of the classical failure factors of ERP implementations found in academic references, text books and research articles (Binji, P. et al.; 1999; Davenport, T., 1998; Lycett, M. et al. 1999; Schwalbe, L, et al., 2000; Akkerman, H. et al., 2002; Turban, E et al., 2006). Among those identified factors are:

1. **Lack of top management support**

In almost all cases of major ERP implementations, the top management involvement and support is considered a tipping point between potential success and failure. At some point of the BaaN ERP implementation project at GB group, management was not fully involved as they should be. At first, the management looked to the project as externally provisioned services by the contracted firm. This has impeded making radical changes (re-engineering) of the business processes and standards within the group's business practice. Such re-engineering endeavors are the norm for successful ERP solution implementations. It was a surprise, for example, to see the requirements document defining separate non-standard business rules among the group's companies. This is by no means what the top management would have required had they been involved in the requirements definition phase of the implementation. This particular issue has led to the fragmentation of the project to a separate company implementation route, impeding any future integration between the instances within the ERP system. Enforcing the integration at the requirements definition phase would have needed the authority and clout of the top management.

2. **Lack of Project team competence**

In reviewing the project team composition, it was striking to see that the implementer is a start-up firm with no or little prior experiences and/or showcases of similar size projects. Further, although the BaaN contract was signed with a regional software firm affiliated to BaaN, the number of implementations provisioned through this firm was also meager to guarantee the success of such a huge undertaking. The counterpart team of the GB group was relatively small and with little or no previous BaaN or integrated information systems experience. Their training was shortened in favor of early start of the implementation process.

The introduction of BaaN ERP system in the region was also so sparse, implicating the scarcity of experienced professionals in both project management and functional know-how using the BaaN ERP system. This explains why GB group ended up with one single qualified trained staff with no redundancy or back up personnel who can effectively operate and maintain the system. This kind of organizational deficiency placed the group in a liability situation impeded any efforts for professional development plans for the existing scarce professionals, who were denied the opportunity to leaving the operation for extended periods for training.

3. **Lack of project ownership**

Like any project, if there is no champion of the project acting as the owner of it, the chances that the project would succeed are usually slim. The pre-requisite qualification of this owner in the group would be one who is not affiliated to one particular subsidiary – but has the group-wide responsibilities. There was no such candidate in the group with such credentials except the Chairman or the group's CFO (back in 1998); both candidates were not acting as the BaaN project champion. Instead, each company's top manager acted as the champion of his individual company's implementation project, furthering the "silo" approach which was reflected in the architecture the BaaN ended up with.

4. **Lack of Standardization**

The lack of standardization on both levels of data structures and processes among the group's subsidiaries, as postulated earlier entrenched the fragmentation and increased the level of customization of the BaaN system. As Engineer Adel Nassim contends, highlighting the gravity of this customization;

"To show how the extent of customization impeded the success of BaaN, the Indian expert who was on board since 2002 was not able to upgrade the BaaN version from IV to V. The effort to do that would have amounted to man-days equivalent to a fresh BaaN implementation in order to incorporate the number of customizations put in version IV. He was barely able to upgrade the operating system AIX version from 4.1 to 4.3 which did not add much functionality to the system."

Engineer Nabil Naguib, the then database administrator for the currently operating Oracle system provides the following account on the magnitude of the inconsistencies in the databases

Table 5. BaaN master database per each company

Company	Customers	Vendors	Items
GE	29087	2517	121102
EMT	29212	2518	1680
ITAMCO	12857	3069	50278
PRIMA	12317	4421	17736
VCI	41	4385	5144
Interland	2808	4421	48283
Total BAAN	86322	21331	244223
Oracle	21366	2154	41451
Repeated	64956	19177	202772

as a result of the non-standardization of the group's data structures;

"The biggest task we encountered when we started the migration of the master databases from the BaaN. The amount of inconsistencies found in these disparate databases was humungous. The code duplications and the inconsistencies were of such a magnitude that it took us months to perform data cleansing. As you can see from Table 5, the repeated records in the source data were nearly triple the number of the actual unique codes for the customer data, nearly 9 times for the suppliers and 5 times for the items data. The effort we expended for the cleansing of this data is humongous. This by no means resulted in 100% clean master database, some items in the master files were still need to be cleansed and were left to be individually dealt with as we come across them in the current Oracle operation".

5. Vendor support and partnership

Although at the time of selecting BaaN as the application software for the information systems infrastructure for the group, this ERP package was considered one of the top 3 ERP systems world-wide. However, the selection criteria did not take into consideration other critical factors in the selection such as:

- Vendor support capabilities in the region;
- Availability of professional skilled staff in the local market;
- The size of the customer base of the vendor in the region;
- Proximity of the support location.

In this case implementation, it was clear that the above critical factors were not readily fulfilled in a satisfactory way to warrant embracing such a complex undertaking. Time has proved that those factors were detrimental to the success of the group's implementation project.

6. Architecture choice

The high level design for the BaaN implementation, in acknowledging the legal requirements for the separation of the group's companies, developed an architecture based on separate instances. This compromised the fact that these separate entities have common back office business processes. For example, the finance and accounting division operates within the framework of corporate resources allocation which proved impossible to implement in this separate instances architecture. This has led to dividing the finance and accounting function to logically disjoint units - each responsible for an instance (company) without regard to the integration nature of those companies. Had the implementation considered the alternate approach in using the BaaN "group" facility coupled with some re-engineering of the integration requirements to cater for the legal considerations, the project may have had a different fate.

Despite the good intentions of the implementation of the BaaN system during the period from 1998 to 2004, the objectives to gain better control over the disparate databases and their integrity were never accomplished. To the contrary, the openness of the BaaN system resulted in the proliferation of the "Excel" culture, fostering the non-integrity of the group's data resources. Based on the findings of the diagnostic study a recommendation document outlining the strategy for implementing new responsive and integrated information systems infrastructure was presented to the management. The strategy addressed the following six major actions:

1. Phasing out the current BaaN ERP system in favor of implementing a new one that avoids the pitfalls of the previous experiences.
2. Development of a blue print for the communications and network infrastructure to serve the current and future connectivity needs for the operation of the integrated business information systems model. Such a blue print must satisfy the following operational requirements:
   - Redundancy;
   - Reliability;
   - Diversified providers;
   - Ample and extendible capacity;
   - Provides good and solid security and safe communications environment;
   - Use of tie lines, MPLS, WiFi and RF technologies for the connectivity of the geographically dispersed locations;
   - Use of the VPN technologies to facilitate virtual presence and mobile access in a secure manner.
3. Establish modern, versatile and high availability datacenter with back up facilities. Acquire and install new server farm infrastructure that will enable serving both the needs for the new information systems infrastructure and the network and communications infrastructure.
4. Establish highly responsive IT team with in-house capacities for operating and maintaining the trilogy building blocks: the information systems, the network and the server farm;
5. Enforce centralized IT service delivery, centralized IT planning process and standardized data structures, databases and processes.
6. Develop the Standard Operating Procedures and Policies (SOP&P) that govern and ensure adequate IT service level delivery to the group's divisions and units.

It has become apparent that there is a need for something more when it comes to the information system at GB group. On the issue of the appropriate selection of the new ERP system, the Oracle e-Business Suite (EBS) was selected. The rationale for this decision was based on several key factors, many of these mentioned previously in the case analysis as problems during the BaaN implementation. There is a local support base for Oracle in Egypt. There is evidence of sustainability and commitment of Oracle to the Egyptian Market. At the time, the business outlook of Oracle is clearly on the rise. EBS functionalities cover almost every line of the business that GB Group is or will be involved in. Also, there is an availability of knowledge workers trained on the full spectrum of the needed professional activities in support of this project.

Phase 3: Technology Components: The New Oracle ERP Implementation

By April, 2004, RGI – the then holding company of the GB group, signed a contract with Oracle Egypt for the licensing of its EBS software which included the following 18 modules:

- Finance which included the five modules – GL,AR, AP, FA and CM
- Distribution which included the three modules – OM, PO and INV
- Human Resources which included the modules – HR foundation and Payroll
- Systems utilities which included – Web Discoverer and Alert manager

- Manufacturing modules which included the six modules – BOM, QC, WIP, MRP, CRP and MPS

Concurrently, RGI invited prospective local Oracle implementation firms to bid for the implementation of the EBS in the group. Two finalists were identified by April, 2004. The implementation contract was finally awarded to the Egyptian implementer CITE company. Their presented technical solution acknowledged the architecture of one implementation instance and not separate instances for the group's companies. This particular concept was the tipping factor in the evaluation of the offered proposals and the selection process of the implementer. Such a solution was thought to guarantee the enforcement of standardization of data structures, database design and the business processes across the entire group's companies.

The implementation plan called for phasing the process into two stages. The first would include the finance, distribution, human resources and the utilities modules; while the second stage will complete the implementation of the manufacturing modules.

In deciding on the operating system platform, use of UNIX operating system as the system layer for the Oracle application was first recommended. This decision was in-line with the best-practice adopted by many organizations for operating Oracle system. However, this recommendation was contested by the management. Among the reasons for this, the not-very satisfactory experience with IBM's AIX support during the BaaN operation. It was subsequently agreed to adopt the option of using Windows 2000 server operating system platform with the caveat of mitigating security vulnerabilities of the Windows platform by deploying adequate security infrastructure. At that time, the 64-bit Windows server architecture was not yet released by Microsoft. This particular issue will be shown to negatively impact the sustainability of the Oracle ERP operation using this platform.

On the organizational level, the consultant team, together with the senior management opted for recruiting an Oracle Project team. The proposed composition of this team is shown in Figure 3. Within few months of setting up the project, the project team positions were duly filled with appropriate professional staff, some were drawn from the existing IT organization, others hired from the market.

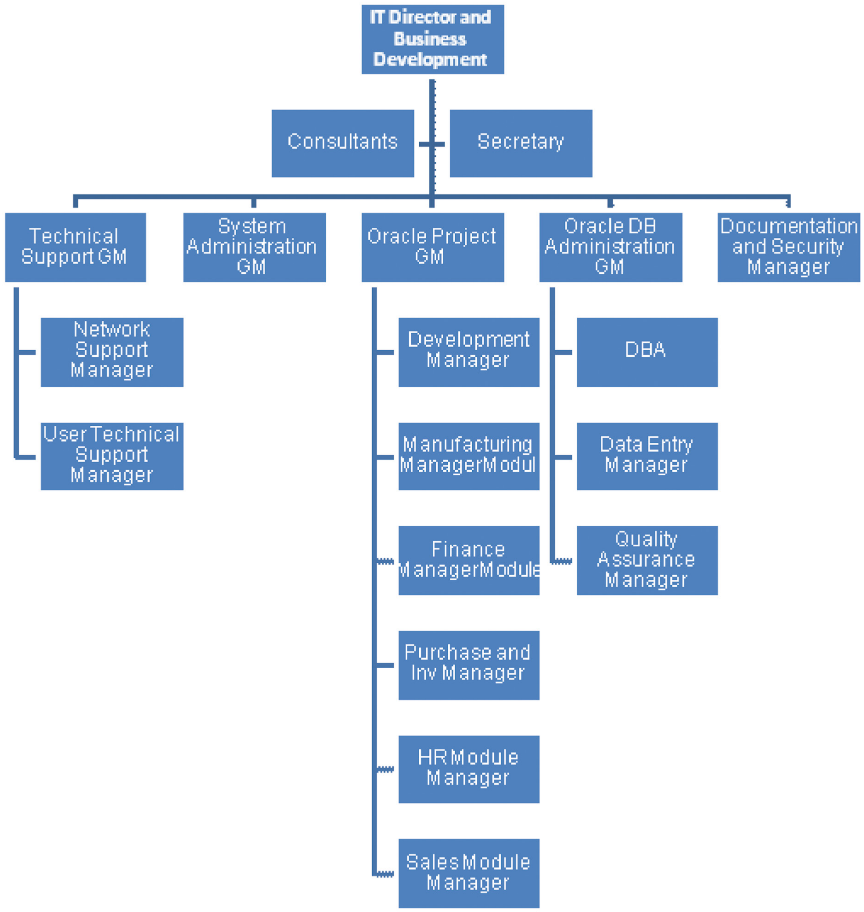
An Oracle Steering Committee (OSC) was also formed. The OSC was to meet on a regular basis to set the project execution strategies and policies, and review the progress presented by the Oracle PM and recommend the necessary actions to deal with any problems or delays in the project time line. The committee composition is as follows:

- The Vice-Chairman for support functions
- Chair of the OSC
- Representatives from the implementation firm CITE
- Representatives from Oracle Egypt
- The IT and Business Development director
- Oracle Project manager from GB
- Oracle DBA from GB
- An external IT Consultant (the author)

Management and Organizational Concerns

The original Oracle project implementation plan stipulated, with the consent of the project steering committee, the conclusion of stage-1 in nine months starting by May 1st, 2004. However, with the unfortunate event of fire breaking in one of the major buildings of the Prima manufacturing company in Abu-Rawash site in December 2004. This resulted in the relocation of the finance and accounting division from this site to the GE company site, some 40 kilometers away from the group's headquarters in Abu-Rawash. Besides the relocation, it was impossible to keep the senior management focused on the project schedules and

Figure 3. The IT organization chart for the GB group during the implementation phase



commitment to the Oracle project in such disaster and damage-control period. The "Go Live" milestone date was subsequently moved to June 2005.

Meanwhile, the project team successfully configured the hardware and the system software infrastructure and readied the platform for the operation of the Oracle application in early 2005.

Soon after, the BaaN databases conversion activities stage started. In this stage, activities such as data cleansing and standardization of data structures, were completed and the databases were uploaded into the readied platform thereafter. Testing, end-user training and go-live for stage one of Oracle project started in early November 2005.

During the live operation of the EBS on the Windows platform, the project team encountered many difficulties and obstacles. The most serious of those was related to the in-adequacy of the WINTEL platform to withstand both the database size and the number of concurrent users using the Oracle application. Although the sizing of the platform was performed collaboratively by both Oracle and HP, the vendors for both the application and the hardware respectively, this glitch persisted. The issue was raised to the top severity level at Oracle's Technical Assistance Request group (TAR) who recommended many solutions and fixes with no avail. It was finally affirmed by both Oracle and Microsoft that the problems encountered were primarily due to the 32 bit addressing limitations inherent in the 32-bit Windows server operating system. The implications of this issue were debated in the project steering committee and a decision was reached to replace the HP/Windows 2000 platform by IBM/UNIX platform. As a consequence, two IBM 570 pSeries machines were procured running AIX operating system. The platform was operationally ready by October 2006. Subsequently, the EBS system was successfully migrated to the new platform and operated by October 2006.

Due to the impending IPO date, the BaaN operation was extended to the end of 2006 and Oracle was placed on the backburner until then. The "Go Live" date was set for December 3,1 2006. This of course had a negative impact on the process of institution and adoption of the new Oracle EBS system. It only meant that user training sessions would have to be repeated and online transaction data entry would cease in favour of BaaN entry.

The go live plan called for populating the balances in all data structures by the 31st of December, 2006. With concerted efforts between the various stakeholders, the system operation started as planned on January 1st, 2007, marking the real start of the third phase of IT/IS deployment trajectory for the GB group.

CURRENT CHALLENGES FACING THE ORGANIZATION

January 1st, 2007 brought about the date that the BaaN system operation was plugged off, and Oracle EBS was put into gear. The IT division at RGI offices was overwhelmingly busy with maintaining smooth and painless transition from the old BaaN culture to the new environment. This was not at all their first encounter with the Oracle EBS live operation. As described before, for a short period of time during 2006, the system operation was in full force – which represented excellent rehearsal of what to be expected after the "Big Bang" of day one in the live of Oracle at GB group. During this period, the team tackled most of the technical glitches related to either the hardware, operating system or the application. It was not therefore a surprise to observe no panic or "fire fighting" reactions in the early weeks after the second EBS "go live" kick off. Nevertheless, change management procedures were noticed to be overly overseen. The after effects of the "cultural changes" started to pile up. The old "Excel" model began to slowly creep back into the company's business process operations. This was manifested by the overwhelming requests made by the users from the business area coordinators for the development of "Excel-like" discoverer reports (reports that are tailor made using the Discoverer tool of Oracle). By midyear 2007, there were over 400 discoverer reports, 90 percent of them mimicked the Excel tabulation. In their pursuit to appease and entice the end-user, the Oracle team submitted to these requests without questioning their effectiveness and /or use that may be accomplished through otherwise standard Oracle built-in reports.

As we mentioned in the beginning of this Case, year 2007 witnessed the start of the preparation for IPO offering under the group's holding company GB Auto. This has created a heavy load on the finance and accounting division, which demanded responsive and agile information system infrastructure. With the limited experienced professional staff in the IT division, handling of those excessive user requests was less than satisfactory from the user's point of view. Coupled with the natural resistance to use the new system in timely manner – the Oracle implementation was perceived by the user as yet another IT failure. At the transaction processing level (TPS), populating the database was in some business areas not performed in on-line, real-time mode. In other instances, entry to the TPS was in error and/or incomplete due to use of untrained employees. The result of such inaccuracies yielded database that does not reflect the real world and a user community without confidence in the Oracle system outputs.

On the management control level, with little or no training in using the Oracle standard reporting system; managers, whether in the middle or top levels, depended on their subordinates' reporting. Those subordinates were compelled to respond through the only way they seem familiar with, downloading data from Oracle and use "Excel" to formulate the required reports. This behavior has proliferated right from the first level management to the most senior level, fostering the perception of the inability of Oracle system to respond to their information needs.

The Oracle post-implementation review conducted by the author and the technical staff suggested that the newly introduced ERP system is passing through critical stage in its adoption path. From an IT planning and control perspectives, it appeared that all the technical ingredients necessary for successful delivery were in place – providing the illusion to the IT management and staff that they have a successful implementation. Blaming the business units/personnel and management of not being receptive to the needed changes. Claiming resisting control and fighting the cause of adopting Oracle within the fabric of the GB group information flow and decision making pattern. On the other hand, focus interviews and direct observations of the various levels of user groups, we found a general perception that the current provisioned Oracle services are not in-line with their needs, neither on the transactional nor on the management reporting levels.

SOLUTIONS AND RECOMMENDATIONS

Following the GB Auto IPO, the CEO together with the BOD members realized that the group's successful growth resulted in the expansion of the organization with a mixture of decentralized management functions at the Business Unit level, while other management functions remaining centralized. Furthermore, the current organization structure showed a relatively high level of vertical layers within each major function, with no clear and effective horizontal communications. This structure led to delays in decision making and impeded progress of work flow, as well as creating conflicts that require top management involvement.

An organizational review has been contemplated to address these issues. Several potential management consulting firms were solicited to bid for this task in early October 2007. The request for proposal objectives included:

1. Working with the GB group's management team to develop the most effective responsive (and lean while capable of continued growth) type of organizational structure, with clear definitions of:
    - Centralized functions versus functions at the Business Unit level
    - Structure of each major areas or function, including definitions on the scope of responsibilities and attributions for each key position
2. Optimization of the operational and administrative processes under the new organization, and incorporating the use of the IT systems to the extent possible.
3. An evaluation of the existing human talent, and a needs assessment in view of the current and future needs of management and professional talent

GB Group intended that the review will focus on those areas that will ensure the effective management of the group and its varied lines of businesses. In January 2008, GB contracted the selected management consulting firm from the three bidders and work on the project started soon after. Among the objectives of the consultant's tasks are the following:

1. **Optimizing processes across the organization**

For a growing organization being able to scale processes and systems is essential to avoid constant reengineering. Also leveraging benefits of scale of similar activities across business units makes managing a growing business simpler. Comparison with standard processes and review of the current processes against a set of best practice industry standard processes will help assess where the biggest gaps are and the advantages of moving towards standardized processes and if needed redesign those non-standardized ones.

2. **Developing effective organization structures and role accountabilities**

All employees who work with a process should be clear on what their accountabilities and responsibilities are. By making this clear and structuring the organization effectively it is possible to reduce the conflicts and confusion between different parts of the organization.

Meanwhile, with the arrival of a newly appointed Chief Finance Officer (CFO) in the senior management level for the group in early 2008, he soon observed that the IT team from one side was accusing the user community of not being receptive to the change. On the other side, users grew more and more resistant to Oracle system adoption, perceiving that Oracle will have to work in the way they are accustomed to and finally consider the Oracle application as a failure since it cannot be customized to their liking.

With ample knowledge of the central role of ERP systems and the implications of its adoption in business organizations, the CFO established an independent task "Oracle Facilitation and Coordination" (OFC) division. The objectives and scope of work of this division were set as follows:

- Review the effectiveness and current use of the current Oracle processes, modules and reports, inform users of all that is available and assist on any alterations or additions needed to the existing Oracle set up in order to operate effectively
- Analyze the current business processes in the various business areas, assess their correctness, validity and appropriateness always keeping in mind operational controls and if needed, redesign those processes to be streamlined with the Oracle configuration
- Document and understand those reports required at the business units and not currently available in Oracle and arrange for them to be included in Oracle by the Oracle system maintainers
- Document reference policies and procedures (P&P) for all business processes in the business functions that have been analyzed as defined in the above points and then monitor the operation on an ongoing basis
- Assess for development any additional KPIs for all management levels in addition to those already existing as well as those in the Oracle DBI module. Coordinate with the IT division for any implementation of business intelligence tools to secure rational approach in managing the group's business resources and operations
- Participate in the implementation of new Oracle modules and assisting in communicating and training end-users
- Conduct on-going training on existing modules to all personnel involved in the use of the developed tools and redesigned operations
- Advise IT division on user requirements and security issues

The scope of work for the new task force would span the entire group's business activities. Work to be done includes:

- Analyze, design and /or redesign the business processes in all the group's line of business
- Analyze, document and evaluate the current configuration of business process operations within the Oracle system and establish gaps between the business needs and the services provided by the Oracle system
- Develop the necessary framework needed for the institution of the Oracle in all business units
- In collaboration with the IT teams and the business unit representatives, develop, institute and operate the business intelligence facilities, either within Oracle or otherwise to maximize the benefits from using the centralized, integrated databases. In collaboration with the group's management, establish the mostly needed KPIs to manage the vast resources of the group
- Provision of on-going monitoring mechanism, overseeing the utilization of, and adherence to Oracle's P&Ps
- Ensure smooth integration of the various activities of the group and institution of any new activities as a result of expansion in new line of business, within Oracle system operation

To that end, efforts of the OFC division team would be aligned with the organization-wide process reengineering initiative to be executed by the management consulting firm, in a concerted effort to pave the way for the institution of the Oracle EBS.

REFERENCES

Ackerman, H., & Van Heden, K. (2002). Vicious and virtuous cycles in ERP implementation: A case Study of interrelations between critical success factors. *European Journal of Information Systems*, *11*, 35–46. doi:10.1057/palgrave/ejis/3000418

Bingi, P. (1999). Critical issues affecting an ERP implementation. *Information Systems Management*, *16*, 7–14. doi:10.1201/1078/43197.16.3.19990601/31310.2

Brazel, J., & Dang, L. (2008). The effect of ERP system implementations on the management of earnings and earnings release dates. *Journal of Information Systems*, *22*(2), 1–21. doi:10.2308/jis.2008.22.2.1

Ciborra, C. (2000). *From control to drift*. Oxford, UK: Oxford University Press.

Ciborra, C., & Hanseth, O. (1998). From tool to Gestell. Agendas for managing information infrastructures. *Information Technology & People*, *11*(4), 305–327. doi:10.1108/09593849810246129

Davenport, T. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*, *76*(4), 121–131.

DeLone, W., & McLean, E. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95. doi:10.1287/isre.3.1.60

DeLone, W., & McLean, E. (2003). The DeLone and McLean model of Information Systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30.

Fan, M. (2000). The adoption and design methodologies of component-based enterprise systems. *European Journal of Information Systems*, *9*, 25–35. doi:10.1057/palgrave.ejis.3000343

Fleck, J. (1994). Learning by trying: The implementation of configurational technology. *Research Policy*, *23*, 637–652. doi:10.1016/0048-7333(94)90014-0

Francalanci, C. (2001). Predicting the implementation effort of ERP projects: Empirical evidences on SAP R/3. *Journal of Information Technology*, *16*(1), 33–48. doi:10.1080/02683960010035943

Fryling, M. (2005). *ERP implementation dynamics*. Information Science and Policy University at Albany. State University of New York.

Gable, G., et al. (2002). Enterprise resources planning systems impacts: A Delphi study of Australian public sector. *Proceedings of the sixth Australasian Conference of Information Systems (PACIS 2002),* 2-4 September 2002, Tokyo, Japan.

Gable, G., et al. (2003, December). *Enterprise systems success: A measurement model*. Paper presented at twenty-fourth International Conference on Information Systems, 14-17, Seattle, USA.

Ifinedo, P. (2006a). Extending the Gable et al. enterprise systems success measurement model: A preliminary study. *Journal of Information Technology Management, 17*(1), 14-33.

Ifinedo, P. (2007). Investigating the relationships among ERP systems success dimensions: A structural equation model. *Issues in Information Systems*, *8*(2).

Ifinedo, P., & Nahar, N. (2007). ERP system success: An empirical analysis of how two organizational stakeholder groups prioritize and evaluate relevant measures. *Enterprise Information Systems*, *1*, 25–48. doi:10.1080/17517570601088539

Krumbholz, M. (2000). Implementing enterprise resource planning packages in different corporate and national cultures. *Journal of Information Technology*, *15*(4), 267–280. doi:10.1080/02683960010008962

Kumar, K., & Van Hillegersberg, J. (2000). ERP: Experience and evolution. *Communications of the ACM*, *43*, 23–26.

Kumar, V. (2001). An investigation of critical management issues in ERP implementation: Empirical evidences from Canadian organizations. *Technovation*, *23*(10).

Lee, Z., & Lee, J. (2000). An ERP implementation study for a knowledge transfer perspective. *Journal of Information Technology*, *15*(4), 281–288. doi:10.1080/02683960010009060

Light, B. (2001). The maintenance implications of the customization of ERP software. *Journal of Software Maintenance and Evolution: Research and Practice*, *13*(6), 415–429. doi:10.1002/smr.240

Lycett, M., & Paul, R. J. (1999). Information Systems development: A perspective on the challenge of evolutionary complexity. *European Journal of Information Systems*, *8*, 127–135. doi:10.1057/palgrave.ejis.3000315

Madapusi, A., & Kuo, C. (2007). Assessing data and information quality in ERP systems. *Proceedings of the Decision Sciences Institute Annual Meeting,* Arizona.

Markus, M., & Tanis, C. (2001b). The enterprise system experience – from adoption to success. In Zmud, R. (Ed.), *Framing the domains of IT management: Projecting the future through the past* (pp. 173–207). Cincinnati, OH: Pinnflex Educational Resources.

Mensching, J., & Corbitt, G. (2004). EPR data archiving- a critical analysis. *Journal of Enterprise Information Management*, *17*(2), 131–141. doi:10.1108/17410390410518772

Moon, Y. B. (2007). Enterprise Resource Planning (ERP): A review of the literature. *International Journal of Management and Enterprise Development*, *4*(3), 235–264. doi:10.1504/IJMED.2007.012679

Nah, F. H. (2001a). Characteristics of ERP software maintenance: A multi-cause study. *Journal of Software Maintenance and Evolution: Research and Practice*, *13*(6), 339–414. doi:10.1002/smr.239

Nicolaou, I., & Bhattacharya, S. (2007). Organizational performance effects of ERP systems usage: The impact of post-implementation changes. *International Journal of Accounting Information Systems*, *7*(1), 18–35. doi:10.1016/j.accinf.2005.12.002

Nolan, R. (1979). Managing the crisis in data processing. *Harvard Business Review*, *57*(2), 115–126.

Parr, A., & Schanks, G. (2000). A model of ERP project implementation. *Journal of Information Technology*, *15*(4), 289–304. doi:10.1080/02683960010009051

Sawyer, S., & Southwick, R. (2002). Temporal issues in information and communication technology enabled organizational change. *The Information Society*, *18*, 263–280. doi:10.1080/01972240290075110

Schwalbe, K. (2000). *Information Technology project management*. Cambridge, MA: Course Technology.

Seddon, P. B. (1997). A re-specification and extension of the DeLone and McLean model of IS success. *Information Systems Research*, *18*(3), 240–253. doi:10.1287/isre.8.3.240

Sedera, D., et al. (2002). Enterprise resource planning systems impacts: A Delphi study of Australian public sector organizations. In *Proceedings of the Pacific Asia Conference on Information Systems (PACIS)* (pp. 584-600). Tokyo, Japan.

Sedera, D., et al. (2004). Measuring enterprise systems success: the importance of a multiple stakeholder perspective. In *Proceedings of the 12th European Conference on Information Systems* (pp. 1-13). Turku, Finland.

Soh, C. (2000). Cultural fits and misfits: Is ERP a universal solution? *Communications of the ACM*, *43*(4), 47–51. doi:10.1145/332051.332070

Spott, D. (2000). Componentizing the enterprise applications packages. *Communications of the ACM*, *43*(4), 63–90. doi:10.1145/332051.332074

Turban, E. (2006). *Information Technology for management: Transforming organizations in the digital economy* (5th ed.). John Wiley & Sons Inc.

Willcocks, L., & Sykes, R. (2000). The role of the CIO and IT function in ERP. *Communications of the ACM*, *43*, 33–38. doi:10.1145/332051.332065

Xu, H. (2002). Data quality issues in implementing an ERP. *Industrial Management & Data Systems*, *102*(1), 47–58. doi:10.1108/02635570210414668

KEY TERMS AND DEFINITIONS

Action Research: Is a process of progressive problem solving led by individuals working with others in teams or as part of a “community of practice” to improve the way they address issues and solve problems. Action research can also be undertaken by larger organizations or institutions, assisted or guided by professional researchers, with the aim of improving their strategies, practices, and knowledge of the environments within which they practice.

Business Process Management: Is a management approach focusing on aligning all aspects of an organization with the wants and needs of the stakeholders. It is a holistic management approach that promotes business effectiveness and efficiency while striving for innovation, flexibility, and integration with technology. If business process management attempts to improve processes continuously, the approach is then described as continuous process improvement (CPI). While, if the approach attempts to revamp the process, then it is often called business process re-engineering (BPR).

Enterprise Resources Planning (ERP): Is an integrated computer-based software system used to manage organizational internal and external resources. It is an architecture which facilitates the flow of information between all business functions within the boundaries of the organization and manages the interfaces with outside stakeholders.

ERP Implementation: To implement ERP systems, companies often seek the help of an ERP vendor or of third-party consulting companies. These firms typically provide three areas of professional services: consulting; customization; and support. The client organization often employs a change program management, business process analysis, and deployment management specialists to align their business requirements with the embedded ERP business processes during implementation

IS Critical Success Factors: Is the term for an element that is necessary for IS project to achieve its mission. They are the critical factors or activities required for ensuring the success of deploying and adopting the IS in the organization. They represent those managerial or enterprise area that must be given special and continual attention to bring about high performance of implemented IS.

IS Success Model: Is a framework and model for measuring the complex dependent variable in IS research. IS Success Model is attributed to DeLone and McLean who originally developed and published the model in 1993 with an update in 2003, which discussed the utility for measuring IS system success within the framework of e-commerce systems.

Longitudinal Case Study: Is a research study methodology that involves repeated observations of the same items over long periods of time. Longitudinal studies are often used in social science research context. The reason for this is that unlike cross-sectional studies, longitudinal studies track the same people, and therefore the differences observed in those people are less likely to be the result of cultural differences across generations. Because of this benefit, longitudinal studies make observing changes more.

Nolan Stage Model: The stages-of-growth model is a theoretical model for the growth of information technology (IT) in a business organization. It was developed by Richard L. Nolan during the 1970s, and published by him in the *Harvard Business Review*

WINTEL Platform: WINTEL is a portmanteau of Windows and Intel. It usually refers to a computer system or the related ecosystem based on an Intel compatible processor and running the Microsoft Windows operating system.

This work was previously published in Cases on Business and Management in the MENA Region: New Trends and Opportunities, edited by El-Khazindar Business Research and Case Center, pp. 177-203, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 52
Matching Manufacturing and Retailing Models in Fashion

Simone Guercini
University of Florence, Italy

ABSTRACT

The aim of this chapter is to examine the interdependencies that have been established with reference to the manufacturer-retailer interaction in textile and apparel (TA). Retailers' strategies seek to reduce the risk of losses from unsold stock, mark-down policies, and stock-outs. These strategies call for manufacturing suppliers to adopt new practices for fulfilling orders flexibility, rapidly, and efficiently. The practices of "lean retailing" imply new manufacturers' strategies, mainly in term of "lean manufacturing." We examine the implications of these processes on the evolution of the relationships between industry and distribution. The chapter addresses the repercussions of the development of lean methods on the development of other formulas having a significant impact on the relationships between industry and distribution, specifically in TA. We then discuss further developments that may be proposed in TA and its channel relationships by shifting from a perspective of supply to one of demand.

INTRODUCTION

This chapter examines the evolution of the relationships between manufacturers and retailers in the fashion system and its implications over time. More precisely, we focus on the manufacturer-retailer interdependencies (Ailawadi et al., 2010) in the textile and apparel (TA) supply chain and distribution channel, which are analyzed as a retail-apparel-textile system (Abernathy et al., 1995). The aim is to provide a picture of these changes and possible future developments[1].

The chapter is divided into two sections followed by conclusions. In the following section we address the repercussions of the development of lean methods on the development of other 'formulas' having a significant impact on the relationships between industry and distribution, specifically in TA. The ability of retail actors to reduce risk and achieve low response times from

DOI: 10.4018/978-1-4666-1945-6.ch052

their manufacturing suppliers has been a powerful driver of change in manufacturer-retailer relationships. In TA the retail strategies of seeking to reduce the risk of unsold stock and losses from breakage call for manufacturing suppliers to adopt new practices for fulfilling orders flexibly, rapidly and efficiently. It is in this context that the concept of 'lean retail' emerged, and hence the practices of 'lean retailing'. Lean retailing is a process that calls for adopting a whole interrelated series of channel practices, beginning at the retail level, with the goal of matching supply and demand, and minimizing the inherent forecasting errors associated with the management of product mixes (Hammond, 1990; Forza & Vinelli, 2000). The introduction of lean retailing began in TA where lead-times were dictated by the two seasonal change-overs (*a-w* autumn winter; *s-s* spring summer), whose importance lay essentially in operational aspects, that is, the lag times between the forwarding of an order and response-delivery (Hunter, 1990). Subsequently, the spread of lean retailing was to place significant stress on all aspects of company programming, including research, production and marketing of the chain-channel. This had effects on decision-making regarding not only product assortments, but also fashion trend offerings and the presentation of collections even at the level of manufacturers, including those in textiles (Agins, 1999). In fact, the adoption of lean retailing operations puts immediate pressure on manufacturers, who find themselves faced with the choice of either holding greater amounts of finished merchandise in reserve in order to meet potential customer demand, with the risk however of winding up with large surplus stock, or innovating production processes to meet the new demands of retailers and thereby reduce their own exposure to the risk of surpluses at the same time.

In the second section we then discuss the implications of adopting lean manufacturing strategies not only in clothing supplier production processes but also at other levels. The requirements of lean retailing process concern not only apparel producers, but textiles manufactures as well. At the same time, lean retailing put pressure on fashion-oriented businesses to adjust the organization of their creative cycle itself in order to satisfy the final demand.

Overall, the chapter highlights the ways in which retailers are able to reduce risk and propose more attractive offerings to consumers. Advances in IT have brought about significant reductions in the cost of information collection, processing and distribution. Already during the 1980s and 1990s, such innovations set the stage for the implementation of distribution strategies aimed at reducing retailers' exposure to the risks associated to the disparate, constantly changing nature of market demands. The ability to use real-time information from points of sales allowed for adjusting product offerings on a daily basis to best meet the level and profile of the effective demand (Stabell & Fjeldstad, 1998). Such strategies aim to reduce inventory levels as a whole. Trends in this direction were already cited in the literature of the 1980s and were destined to increase considerably over the following decade (Tyler, 2008).

It should however be noted that the information gap remains the weak point in both retailers and manufacturers' relationships with the market. Given the current state of affairs, struggling manufacturing companies need to take a more active role in dealing with their demand-chain, especially by anticipating demand, while retail companies has been able to take an active role in the supply chain. A number of success stories may serve as examples: some manufacturers have managed to thrive by establishing tight relationships with retailers who have developed the capacity to give them a competitive edge in this rapidly evolving sector.

An outline of this evolution and some conjectures about future trends are presented in the concluding section. The relevant issues are addressed through a review of the literature and discussion of the author's own research.

Evolution of the Channel and 'Lean' vs. 'Outlet' Formulas

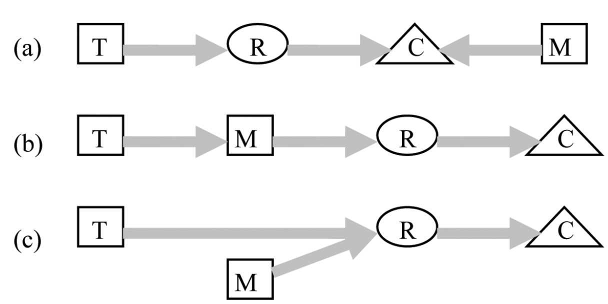
The spread of lean retailing processes over the last twenty years or so serves to highlight a return to the centrality of retail, in which the actors upstream from the chain-channel are in more direct contact with the distribution level. The sequence of traditionally recognized relationships between (1) textile enterprises and clothing enterprises, (2) apparel enterprises and retailers, (3) retailers and consumers has been interrupted, at least in part, by the new centrality of retail. This has translated into the emergence of relationships unaccounted for in traditional models, as retailers may be directly linked as clients to textile firms, and textile firms, in turn, may be direct customers of companies furnishing manufacturing services (Figure 1).

In the model illustrated in Figure 1(c), firms operating at the industry level put their production capacity at the disposal of third parties, and thereby essentially act as contractors. Such enterprises may develop and promote a process of upstream integration of distribution firms, given that that these latter carry out roles typical of industry (sourcing raw materials and semi-finished products, product design, quality control, etc.), which are not carried out by contractors (Weiss, 1958).

In apparel channels characterized by predominately 'independent' relationships, clothing manufacturers interface with the suppliers of semi-finished textile products and retailers essentially through transactions between autonomous agents. Apparel producers operating in channels with largely '*interdependent*' relationships have closer ties with subcontractors and retailers. This results in greater flexibility and a constant exchange of information (Uzzi, 1995). In this case, distribution can furnish information to the manufacturer, with whom it then coordinates actions to carry out merchandising and other marketing activities in their common interests (Sheridan et al., 2006). Such information exchange with the retailer can enable the manufacturing firm to better direct its short-term production, as well as its own design activities. The distribution firm may however decide to take over these critical functions (production planning, design…), provided of course it has the means and resources to do so. Thus, it would not need to develop 'interdependent' channel relationships, but could maintain 'independent' relationships, by which it can make use solely of the manufacturer's production capacity, whose role thus tends towards that of a contractor.

Regarding the last type of distribution channel mentioned, that is, one characterized by '*integrated*' relationships, the roles of both apparel

Figure 1. Supplier-buyer links in the three models considered [Source: our elaboration]



T = textile firm; M = apparel manufacturer; R = apparel retailer; C = consumer:
(a) 'tailoring dominated' model: T(0,1), R(1,1), C(2,0), M(0,1)
(b) 'ready-made clothing' model: T(0,1), M(1,1), R(1,1), C(1,0)
(c) 'retail' dominated model: T(0,1), M(0,1), R(2,1), C(1,0)

producer and retailer are taken on by a single agent, who carries out almost all management functions regarding both manufacturing and distribution operations. Such agents may be firms originating historically in the distribution sector, but who have taken on typically industrial functions, or conversely, they may originally have been involved solely in the manufacturing end and expanded downstream into retailing.

In TA industry-distribution relationships a role of increasing importance is being played by new out-of-town commercial sites called '*outlet centers*'. For manufacturers whose client retailers have adopted lean retailing, such centers can offer an alternative to adopting lean processes to satisfy such clients' needs. Implementing lean retail calls for quick response times on the part of the distribution network. Suppliers that adopt lean manufacturing to satisfy such demands are also able to reduce the negative effects of (1) unsold merchandise, (2) discount sales at cost, and (3) lost sales due to stock breakage, for both themselves and their client distributers. In contrast to this solution, outlet stores, and outlet centers in general, offer a direct sales channel that acts as a 'relief valve' for producers, that is, a means to dispose of excess production, generally through the formula of 'every day low price'. In this sense, the outlet channel can be viewed as an alternative solution to lean manufacturing, in the manufacturer's perspective, though (as we will see) it is potentially complementary to lean retailing in fashion apparel distribution. Despite the higher risk of surplus associated with the higher inventory necessary to ensure rapid service, suppliers of 'lean' distribution firms can maintain their traditional planning because they can later dispose of any such surplus inventory through the direct channel represented by the outlets.

The 'outlet' distribution channel, however, presents at least one disadvantage with respect to the option of responding to lean retailing by implementing lean manufacturing. This is that lean manufacturing allows for the possibility of limiting lost revenues from stock breakage, while this cannot be counted, per se, among the benefits of an outlet channel.

Naturally, we can also envisage types of channels in which sales through outlet stores are combined with lean manufacturing practices, thus regarding the two processes as complementary. In such a scenario, manufacturing enterprises would also undertake a form of integration of the distribution process, albeit through a direct approach, which allows it to dispose of surplus inventory through a dedicated channel. In light of such considerations, it should be no surprise that the outlet channel developed rapidly, in parallel with the spread of lean retailing in the fashion apparel industry – in the United States up until the late 1990s and over the following decade in Europe.

The outlet channel represents a type of 'direct' distribution channel, which is essentially characterized by the presence of an agent, the promoter of the center, whose role is to carry out some of the functions typical of retail. In fact, the promoter may take on certain management functions characteristic of retail activities (hiring personnel, marketing, display management and merchandising), together with other roles that are instead typical of a site promoter, starting with determining tenant-mix strategies and generating traffic by promoting the outlet (Burresi & Guercini, 2003).

The outlet form of direct channel is therefore particularly attractive for operators who are not committed to implementing lean manufacturing in response to the lean retailing adopted downstream by their clients in distribution. The agents that integrate processes of lean manufacturing and lean retailing are therefore by definition those less oriented towards the outlet channel, even though their positioning may in some cases be higher (Figure 2).

The results attainable through integration in channel relationships underlines how tight coordination, prompt information exchange, and risk sharing between industry and distribution are

Figure 2. The relationships between unsold stock management and positioning [Source: our elaboration]

		Positioning	
		lower	higher
Manufacturer unsold sotck management strategy	lean	typical lean retailing and manufacturing	integrated outlet and lean
	outlet	integrated outlet and lean	typical brand manufacturer outlet services

fundamental to providing a timely response to the changes of the final market (Richardson, 1996, p. 410). It seems that such integration need not necessarily be combined with forms of proprietary control. Besides such forms there are others in which integration between distribution and industry may be the result of a bias towards servicing the latter, which is amenable to the development of agreements and alliances with retail firms through channels dominated by 'independent' relationships.

The proliferation of items references (i.e., number of Stock Keeping Units to manage) is a long-term trend in TA, and suggests a series of implications for both producers and retailers, the first amongst which is the greater uncertainty over what product(s) will sell best (Abernathy et al., 1995, p. 192). After all, companies produce a variety of merchandise in order to gain a greater market share (competitive logic), as well as to fulfill the wishes the final consumer (demand logic).

The greater the variety of merchandise, the greater the uncertainty for each single product. In order to furnish more service to clients without winding up with more surplus, retailers must be able to acquire needed goods without delay. Accordingly, this requires that manufacturers be able to fill orders promptly and efficiently. If manufacturers respond simply by increasing their own inventories, they will incur rising costs (especially from unsold stock) and they will, in the middle-to-long term, be less competitive. Sooner or later, at some point client distributors will be driven to seek a more efficient supplier (Blackwell, 1997). Thus, lean retailing inevitably appears to favor the development of lean manufacturing. Adopting lean retailing practices offers retailers a way to capitalize on advances in information technology and to minimize exposure to uncertainties in demand. The demands of clients with such preferences will bring pressure to bear on manufacturing suppliers, engendering substantial changes in the practices of the supply chain, and moreover with consequences on the entire business process (Milgrom & Roberts, 1988). Thus, at one extreme, manufacturers could simply opt to not innovate their internal routines and hold the necessary inventory to service lean retailers, thereby increasing their inventory costs. On the other extreme, manufacturers could upgrade their operations of internal planning and design, provisioning, supply and production, to be able to rapidly respond to changes in demand, and thereby minimize the risks associated with surplus inventory, as much for themselves as for their clients. In an even more developed form, it could be the manufacturers themselves to impel retailers unwilling to follow such an approach to adopt lean forms of retailing, thereby leveraging their own quick response capacity as lean manufacturer to gain a competitive advantage over competitors (Guercini, 2001).

PERSPECTIVES IN THE NEW MANUFACTURER-RETAILER RELATIONSHIPS

As we have seen, the introduction of lean retailing leads to significant transformation in both the standards of service expected and the interactions between industry and distribution, with a shift of manufacturers' investments in distribution centers and service systems towards the retail segment. The development of lean retailing entails the increasing dominance of retailers within channel relationships. The need for suppliers to adapt their behaviors to meet the demands of retailers does not necessarily imply symmetry in the flow, or even sharing, of information, given that retailers simply impose compliance to their standards and are not obliged to share data on demand beyond that information implicit in supply specifications (Hunt & Nevin, 1974).

Distribution practice has also undergone substantial progressive changes in terms of logistics. Traditionally, goods have always been transported from manufacturing plants to warehouses also owned by the manufacturer, whence they would then be transferred to retailers' storehouses. In lean retailing systems, the warehouse has been replaced by automated distribution centers, which interface directly with both industry and distributors (Hoover et al., 2001). Clothing suppliers can set up their distribution operations so as to increase the information links between manufacturing enterprises and retailers, thereby enabling labeling or informing retailers of upcoming deliveries, for example. Thus manufacturers can provide retailers with new services, which may even involve advance price labeling, as well as general merchandising services (Park, 2004).

Innovations in the apparel channel also have effects upstream on the textile market. At times change arrives suddenly, at others it comes more gradually, but it looks a general trend, and it is always the retailers who represent the essential agents of change (Doyle et al., 2006). Retailers call for clothing suppliers to invest in the first, emergence stage of "lean retailing" strategies – a whole series of basic information technologies whose purpose is to meet the requirements of ever-more demanding retail service in terms of short lead-times and rapid fulfilling of orders (Palmer & Markus, 2000; Forza & Vinelli, 2000). Retailers do not care much about how suppliers satisfy such requirements, as long as they do (Geysken et al., 1999). We are now witnessing the ever-increasing emergence of practices that address the problems of interfacing retailers and manufacturers, rather than the purely manufacturing concerns of clothing enterprises. If clothing producers adapt their operations to face the new and rising demands of the channel, their suppliers must also become more prepared to respond with new offerings of finished and semi-finished fabrics, as well as other new products and services (Jones, 2002). Although purchase price has traditionally represented the basic reference element for assessing supply alternatives in the TA industry, the spread of lean retailing strategies has brought other criteria into play, in particular, supply lead time. Lead times are traditionally longer for the delivery of fashion fabrics than for basic ones. In particular, in the formulation of the early 1990's, lean retailing focused mainly on "*basic*" or "*fashion-basic*" goods, rather than "*fashion oriented products*" (Abernathy et al., 1995). The spread of lean retail strategies in the field of fashion becomes evident in the second formulation of such strategies, widely adopted especially in the late 1990's, though their inception had come somewhat earlier [2].

Lean retailing processes, in fact, underwent a second stage, whose peculiar features with respect to the preceding stage include: (1) greater involvement of textile suppliers in the adoption of lean retailing; (2) the spread of lean retailing also to the field of fashion-oriented products, which imposed a reorganization of the stages of planning and designing apparel. A number of other aspects are also interesting in this regard: (a) the retail-apparel-textile channel appears increasing driven by actors who take on the role of lean retailers or, when originating in the manufacturing sector, producers who take on the functions typical of retailers and assist their retail clientele in implementing lean practices; (b) increasing interdependencies arise within the channel, and are linked to the processes of organizational integration, which also promote processes of vertical integration in response to the increasing need for coordination; (c) there is an increasing service orientation in the exchanges within the channel; (d) although the prime movers of change come clearly from the retail level, other actors may also address the problems of interfacing with distribution and consumers; (e) since lean retail requires a "quick response" from the supplier, when dealing with fashion and fashion-basic products, local suppliers may enjoy a relative advantage due to their shorter reaction times (Christopher et al., 2004).

New information technologies, new production process management, the evolution of production costs abroad and new organizational solutions are all aspects that have contributed to transforming the relationships between distribution and industry in TA (Palmer & Markus, 2000). It seems that the close communication within such relationships have come to substitute for inventory reserves (Milgrom & Roberts, 1988). The emergence of lean retailing processes has progressively revolutionized the relationships between distribution and industry in TA. Within the broader context of modern clothing distribution, such revolution may be invoked to partly explain the central importance attained by retailing among the various roles in the TA supply channel. Lean retailers differ from traditional clothing businesses in their greater exploitation of information technologies, an aspect that especially characterizes the first stage of lean retailing's development, and which has brought about a change in the demands on manufacturing suppliers for speed and reliability, thereby driving them to adopt lean manufacturing practices. The second stage of development of lean retailing therefore led to the involvement of other components of the value chain, generally within the manufacturing sector, and particularly affecting the areas of planning and design (Jackson, 2001; Sheridan et al., 2006). The consequent initial increase in the number of seasonal collections (from two up to four or even six) was quickly followed by a veritable explosion, even as far as the creation of weekly collections in fast fashion enterprises. Various critical factors account for the success of seasonal fashion-oriented products (Mattila et al., 2002), all linked to sourcing: *forecast accuracy*; *process lead-time*; *offshore-local sourcing mix*; and *up-front/replenishment buying mix*. In lean retailing forecast accuracy is less critical, because what is produced is tightly linked to the demands of direct customers (Mattila et al., 2002, p. 343).

Brand manufacturers who work with lean distribution clients may find themselves having to manage potentially unsold stock, if they themselves do not "go lean" in their own manufacturing processes, that is to say, adopt *lean manufacturing.* In the event that they do not, then the organization of the supply chain / distribution channel will see to it that that any surplus stock is essentially left with manufacturers. Another way that this can happen is when the industry is integrated at the retail level, which however does not practice lean retailing. Thus, the reserve stock is functional – it is necessary in order to fully serve the sales sector and avoid stock-outs. These two situations are illustrated by "c" and "d2" in Figure 3. In both cases industrial producers can manage these reserves through a direct outlet channel, offering "out of season" and/or "out of standard" products. It is interesting to note how lean manufacturers who interface closely with lean retailers, by limiting the amount of products released into the retail channel, should not need a presence in such outlet channels.

The relationships between industry and distribution in TA can be analyzed by comparing manufacturer management of the "*supply chain*" and retailer management of the "*demand chain*"[3]. As already seen, from the 1980's until the 1990's

Figure 3. Lean strategies in TA manufacturing and retail [Source: authors' processing]

		Retailing: lean	Retailing: non-lean
Manufacturing	lean	(a) limited reserves in channel	(b) retailers hold reserves
	non-lean	(c) manufacturer holds reserves	(d1) retailers hold reserves, or (d2) retail-integrated manufacturers hold reserves

new technologies and new strategies and organizational models were progressively adopted to reduce lead times and limit surpluses due to market forecasting errors. This process has produced *quick response* times for lean retailer clients, but should per se be distinguished from the processes involved in *quick fashion* (Guercini, 2001, p. 70). This latter development represents a formula aiming not only to reduce leads times, but also to adjust organization of the creative cycle itself as a function of the requirements of final demand. In traditional "programmed" production management systems, the fashion chain leading up to a new season calls for an 18-month cycle (Irpet-Textrends, 1991). For instance, during the 1980's the average lead time in the clothing industry, from raw material to final consumer, has been estimated at about 65-70 weeks, of which only 11 were for manufacturing, while nearly 40 were taken up by the processes of wholesaling and logistics. In recent years a rather general rule appears to have defined a 12-month cycle (thus 15-20 weeks less than in the 1980's). The main benefit of this reduction in lead times has been to minimize forecasting errors and their effects. It has been estimated that, while traditional times can lead to forecasting errors of about 40% of overall production, reducing lead times from 12 to 9 months can cut this margin of error by half (Jacobs, 2006, p. 85). Further reductions in lead-times can bring about progressive reductions in forecast errors, with consequent improvements in the overall balance sheets of channel enterprises. It has been estimated that the pre-season forecasts currently suffer from an error of about 10% (Lowson et al., 1999, p. 44, p. 93 ff.).

In this regard, fashion designers conceive of new trends about one year before the target season. However, the decisions regarding semifinished textiles, beginning with the fabrics to employ, are made months before (Jackson, 2001, p. 127 ff). Nowadays, many successful firms have reduced lead times so drastically that they are no longer tied in the same way to the seasonal production management model. Vertically integrated firms such as Inditex-Zara and H&M can complete the entire cycle from design to distribution in only three to eight weeks, thereby effecting substantial changes in operating relationships within the channel.

The development of lean retailing aims to limit lead times and thereby avoid errors in forecasting and, consequently, the associated higher costs / lower profits. Two parts of the TA value chain can be distinguished: the "*supply oriented*" and the "*demand oriented*", as per the well-known scheme proposed by Porter (1985, p. 35). This chain is normally represented with horizontal left-to-right arrows, which indicate the flow from raw materials to final consumers. Thus, we speak of "*demand and supply chain management*", in which the flow of products are oriented in one direction (supply), and the information relevant to decision-making in the other (demand). According to Jacobs (2006, pp. 86-87), some components of the value chain can be represented as flowing from right to left in the "demand" perspective, in which the position of end consumers on the right-most side emphasizes their determining role in orienting and driving enterprise activities in the chain-channel. Because suppliers are clearly crucial to important segments of clothing firms' value system, reducing lead times in the supply chain requires a concerted effort with these partners. Suppliers' orientation or stance with respect to their specific "*demand chain*" is clearly important in this process. In fact, there are two ways suppliers may define such stance. The first way for suppliers to take such a demand stance is to assume that the requisites expressed by retailers is the true expression of demand. In this case, the directly relevant agents of demand for the manufacturing firms are their immediate retail clients. The second approach instead takes the demand stance in perhaps the strictest sense of the term, by equating it directly with end consumption. Consequently, cooperation with retail occurs through different means. In fact, in this case manufacturers may seek to integrate vertically, taking on distribution functions in an effort to bring themselves closer to end consumers.

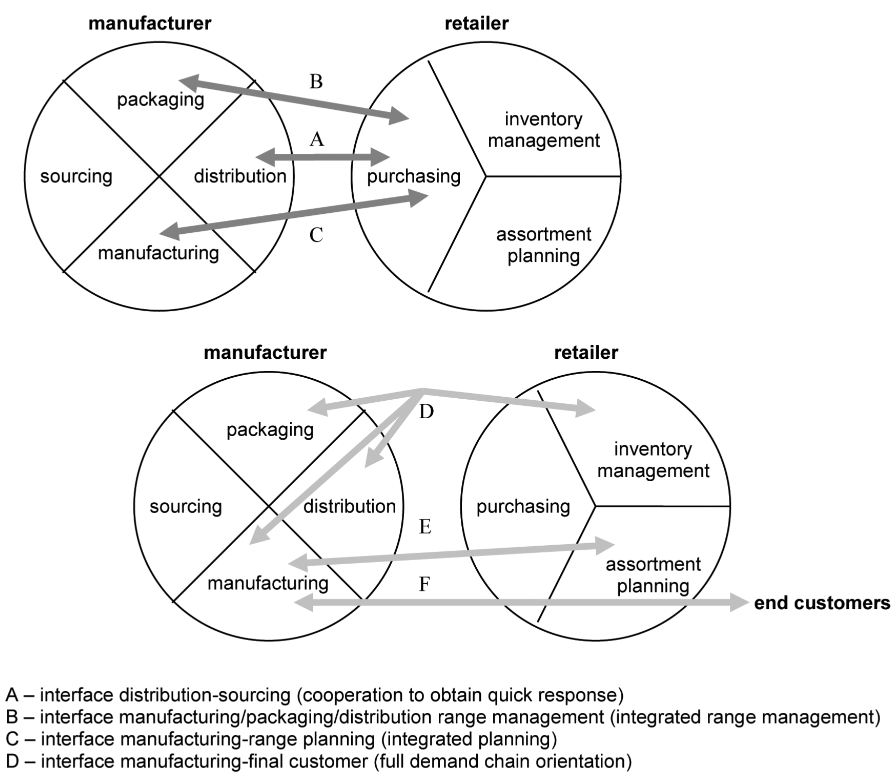
Nowadays, the fact is that for manufacturers the demand side of industry-distribution relationships is still generally represented by retailers, who must plan their own product mix ("what" to sell), manage their inventory ("how much" of each article and SKUs) and consequently decide on purchases ("where" and "who from"). Retailers make such decisions based on their knowledge and assessments of the needs and desires of consumers. Such assessments may however suffer from significant shortcomings, especially considering, as highlighted by some recent studies, that more than half of European fashion consumers fail to find what they are looking for at sales points (Stockert, 2004). The logical conclusion, already drawn by various authors, is that the state of information on consumption remains the "Achilles' heel" of TA chain-channel management. The supply side of industry-distribution relationships is clearly of relevance to retailers, given that they must interface with manufacturers who combine or collect the inputs for production (sourcing), and must organize its various stages, including packaging and delivery of products to the distribution channel.

Analyzing the supply chain and the demand chain is also of considerable importance to the organization of integrated enterprises, as they do not necessarily require distinct, recognizable agents. When such vertical integration has not been implemented, it falls to the manufacturer to study the demand chain, just as it is up to the retailer to examine the supply chain, thereby giving rise to a distinction based on which the links between the two perspectives can be discussed with greater precision.

On the *demand side* the supply chain can be managed through purchasing policy, by seeking different types of links through which to interface with different areas in the typical supply chain of manufacturers (Hoover et al., 2001, pp. 74-76). These are represented by "distribution" (A), "packaging" (B) and production in its narrowest sense, i.e., "manufacturing" (C). Retailers can in fact look for contacts with manufacturers' distribution divisions, with the aim of ensuring rapid delivery, as the supplier has probably invested in substantial amounts of inventory (scheme "A" in Figure 4). In the second approach mentioned, retailers look for contacts with packagers and request assembled products from clothing suppliers in response to precise consumer demands. Such requests may include logistics management, dedicated packaging, adjustments to product specifications and/or private labeling ("B" in Figure 4). Lastly, in the third approach, retailers interface with suppliers' manufacturing processes, from whom they can order custom production (*manufacture-to-order*), though this comes at the cost of lower efficiency and longer delivery times. Such solutions moreover generally involve accepting higher supply prices ("C" in Figure 4). Retailers therefore come to manage these alternative sourcing methods by operating on the demand side and carrying out "*supply chain management*". On the industry side, manufacturers must cope with the requisites and constraints of production technologies. Moreover, they are often responsible for managing the more well-established, if not traditional, requirements of inter-organizational relationships along the TA chain / channel. Often producers who avoid such role, do so by taking on a more demand side approach, that is, they manage process so as to achieve high levels of distribution sell-out by favoring retail activities over direct relationships with final consumption.

On the *supply side*, industry can implement various different policies for managing the demand chain, thereby adopting so-called "*demand chain management*" (Baker, 2004). Four different forms can be distinguished (Hoover et al., 2001, pp. 76-80). The first form, interfacing producers' distribution functions, essentially involves retailer purchasing. Such an approach to demand chain management is a mirror image of the first of the three types of supply chain management cited with regard to retailers ("A" in Figure 4). This first case therefore represents a form of cooperation that can provide suppliers with a quick response strategy to satisfy the requirements of lean retailers (Iyer & Bergen, 1997). A second form of demand chain management by manufacturers calls for joint effort in the operations of manufacturing, packaging and distribution, which may need to interface administratively with purchasing, but which, in any event, seeks to coordinate with retailer clients' inventory management,

Figure 4. Alternative relationships between manufacturer (supply perspective) and retailer (demand perspective) in TA [Source: authors' processing]



thereby providing savings in inventory costs ("D" in Figure 4).

A third type of demand chain management involves establishing relationships between producers' manufacturing divisions and retailer clients' inventory management. In order to collaborate in defining retailers' inventories (including merchandise assortments), manufacturers must, together with their direct clients, try to understand the consumer demand categories that both are seeking to satisfy ("E" in Figure 4). The fourth and last type of demand chain management presents some particular features. Here, manufacturers organize their demand chain management based not on their retail clients, but rather directly on indications from final consumers ("F" in Figure 4).

This last form of demand chain management appears particularly interesting. It has been likened to the strategies used to overcome retailing and achieve mass customization in other sectors, such as the case of Dell in the personal computer industry (Kerin et al., 2009). Such approach to the demand chain may be of as much interest to *manufacturer-retailers* (producer integrated in distribution) as to *retailer-manufacturers* (distributor integrated in production). This latter presumes an inverted perspective of the path: retailers try to achieve a more direct link between manufacturing and final consumers through their own distribution network. In both cases, the relationship between industry and distribution in TA merges the various resources in an integrated fashion to guarantee more direct contact between manufacturing processes and the downstream activities of the demand chain [4].

In such a scenario, industry may also seek to set up new services of interest mainly to the retailer, so as to build up dynamic cooperation through which to improve economic results of equal benefit to both. In the event retailers press excessively for price cuts, the manufacturer may try to establish direct relationships with consumers, following the approach outlined in the fourth and last form of demand management described above. Alternatively, it is sometimes retailers themselves, who by their close participation in typically manufacturing functions, take on the distinctive features of retailer-manufacturers (or industrial retailers), and adopt the fourth approach to demand chain management described, in order to respond more fully to the consumer demands they manage to perceive. In any event, both retailer-manufacturers and manufacturer-retailers come to address the logic of both the supply chain and the demand chain in an at least partly integrated fashion within their enterprises. The understanding of consumer demand that can be gained solely from the information collected by conducting retail activities cannot however account for the so-called "unrealized sales" consequent to consumer dissatisfaction with the assortment offered. The traditional functions of the middleman do not include the developing of close contacts between manufacturers' offerings and final demand. Managing the demand chain does not simply represent the mirror image of supply chain management by the retailer, given the important role that this latter's orientation to final demand may have. This is an aspect that still requires further development, despite the meaningful impetus towards the vertical integration of fashion enterprises. Integration must go beyond the simple, partial forms that developed in the early stages of lean retailing, and broaden vertically to the include interfacing with consumers downstream and with the design processes upstream.

With more complete vertical integration the drive to cooperation among actors within the same channel gives way to a greater impact of competition between channels involving vertically integrated enterprises. Even at the retail level, there seems to be little understanding of the causes of unrealized sales to actual customers (not to mention prospective clients), and to this end information technologies are of little use without the ability to relate to customers, draw nearer to them, gain greater understanding and learn through a relationship with them. The fourth type of supply chain management ("D" in Figure 4) therefore has the potential to bring about more radical innovation – not only in end consumer products and services, but also in the relationships between industry and distribution – than the two stages of development of lean retailing (Jacobs, 2006, p. 92). At the same time, there has been a shift from greater *vertical cooperation* toward one marked essentially by *horizontal competition.* This suggests that the evolution of industry-distribution relationships in TA is potentially destined to devote greater attention to the acquisition of knowledge on the prospective behaviors of the final customers of all businesses, that is, consumers. This comes about in the face of the continued standardization of service centers, as highlighted by the excessive uniformity in the various types of shopping centers, which in turn has stemmed from the pursuit of economy by the large fashion chains and the management and control systems of the promoters that have created them.

One important related aspect concerns the considerable difficulties involved with actually using the data produced by distribution information systems (micro-marketing applied to apparel distribution). Such difficulties stem from the currently limited number of case studies on market segmentation based on the data collected by POS systems. Faced with such limitations in retail information technologies, fashion demand seems ever more difficult to predict, especially in light of the long-term increase in its diversity that began in the 1950's and 60's (Abernathy et al., 1995) and which has accelerated considerably in recent years. Cooperation between retailer and manufacturer does not appear enough to face such rapid development, either with regard to inventory and selection management alone, or at the level of planning product ranges and manufacturing. Moreover, the potential benefits of cooperation in terms of understanding prospective consumption do not seem extend beyond that which has been defined a generic advantage of two partners may possibly know more that one (Jacobs, 2006, p. 89).

CONCLUSION

This paper has discussed how the evolution of the relationships between industry and distribution in TA seems to follow particular paths that are quite distinct from those developing in other fields of distribution. Such paths are shaped by the peculiarities of the manufacturing and distribution processes in the fashion industry and the TA production cycle. More specifically, what emerges from the analysis are at least three possible evolutionary patterns, which in historical terms define the last three decades, at least. However, such paths of evolution also take on a conceptual quality and therefore may be abstracted from any specific time period.

In a first stage of evolution the relationships between industry and distribution are characterized by the early success of lean retail formulas, which shape the operational processes of industry. This upsets the established order, in which both industry and distribution organize their product offerings (textiles and apparel) based on two seasons (spring-summer and autumn-winter). Thus, with the rise of lean retailing, this first evolutionary stage allows for going beyond the traditional patterns of production cycle times, and brings about profound changes in retail enterprises' operations and their relationships to industry. In this stage, enterprises seek to minimize lead times through the collaboration between industry and distribution made possible by the introduction of technological innovations in information management. This stage reached its peak in the 1980's and was characterized by the development of business enterprises, such as Benetton, which represent models for analysis and emulation at the international level.

During the second stage of evolution of industry-distribution relationships the consequences of the adoption of lean formulas spread to involve industry, particularly in their interior functions upstream of logistic and commercial aspects. Once again, such trends are triggered at the level of retail functions and processes and, unlike the preceding stage, they not only tend to shape the operational activities of supply process management, but come to involve and profoundly modify the very creative processes of industry. Within this framework, some manufacturing processes can be integrated directly into enterprises whose prime concern is distribution. The spread of fast (or quick) fashion processes is symbolic of this stage. Not only are operative lead times reduced to a minimum, but conception and design times come to be defined by ever more frequent offerings as determined by the need to bring to market merchandise in a continuous "state of becoming" and hence able to stimulate demand. In this stage the lead times dictated by two fashion seasons have been definitively superseded by a large number of new offerings brought out at an ever more rapid pace. Firms such as Zara are emblematic of this stage of evolution of the relationships between industry and distribution in TA. Cooperation through vertical relationships gives way to processes of true strategic and operational integration, accompanied by the heightening of horizontal competition. There is a full-blown trend toward market- and service-orientation, with a consequent shift from a supply chain in which the guiding principles for buyers-suppliers are defined by technology (upstream), to one in which the (downstream) processes of distribution and consumption play the central role.

This model seems to have reached full maturity in recent years. Enterprises that have adopted such practices have even achieved better performance than those gained through other strategic groups. Such market orientation has led to improvements in industry response times to customers' actual purchasing behaviors and moreover stimulates consumer interest by intensifying the pace of new, and therefore 'fresher', fashion offerings. Such model seeks to minimize the lag times between conception and production, on the one hand, and better respond to consumer purchasing behavior, on the other. However, it does not address the need for knowledge, for instance, on unrealized sales due to a consumer's failure to find the sought for product in stock. In other words, in both the evolutionary stages described, attention is devoted more to purchasing behavior than to actual consumption. This represents the "Achilles' heel" in the market orientation of TA firms. Both industry and distribution enterprises remain largely committed to marketing offers at an ever accelerating pace, with the aims of stimulating the market and reacting rapidly to its manifestations of interest. Even more so than in other fields, TA is still lacking the market research tools necessary to acquire direct knowledge of consumption. A greater understanding, for instance, of unrealized sales and their relations to what customers look for, could have significant implications for the strategy decisions of enterprises, which over the last years have often focused on points of sales and customer-experience aspects. Such an approach marks the third and last stage of evolution of the relationships between industry and distribution, one which could lead to, or at least offer, alternative prospects to the trends that dominated the earlier stages.

For future research, it will be interesting to find analytical evidence to theoretically verify some critical insights revealed in this paper. For example, for channel coordination issues in different stages, the respective operational benefits and coordination mechanisms can be explored (see Eppen and Iyer 1997, Chen et al. 2006 for more details).

REFERENCES

Abernathy, F. H., Dunlop, J. T., Hammond, J. H., & Weil, D. (1995). The information-integrated channel: A study of the U.S. apparel industry in transition. *Broking Papers on Economic Activity. Microeconomics*, 175-246.

Agins, T. (1999). *The end of fashion. How marketing changed the clothing business forever*. New York, NY: Morton.

Ailawadi, K. L., Bradlow, E. T., Draganska, M., Nijs, V., Roderkerk, R. P., & Sudir, K. (forthcoming). Zhang, J. (2010). Empirical models of manufacturer-retailer interaction: A review and agenda for future research. *Marketing Letters*.

Baker, S. (2004). *Demand-driven manufacturing*. Bromsgrove, UK: Just-style.com.

Blackwell, R. (1997). *From mind to market/Reinventing the retail supply chain*. New York, NY: Harper Business.

Bucklin, L. P. (1970). *Vertical marketing systems*. Glenview, IL: Scott, Foresman & Co.

Burresi, A., & Guercini, S. (2003, November). *Nuovi attori e integrazione di funzioni nel marketing strategico della distribuzione*. Paper presented at the conference Le tendenze del marketing in Europa, Venice, Italy.

Chen, H., Chen, J., & Chen, Y. (2006). A coordination mechanism for a supply chain with demand information updating. *International Journal of Production Economics*, *103*, 347–361. doi:10.1016/j.ijpe.2005.09.002

Christopher, M., Lowson, R., & Peck, H. (2004). Creating agile supply chains in the fashion industry. *International Journal of Retail & Distribution Management*, *32*(8), 367–376. doi:10.1108/09590550410546188

Doyle, S. A., Moore, C. M., & Morgan, L. (2006). Supplier management in fast moving fashion retailing. *Journal of Fashion Marketing and Management*, *10*(3), 272–281. doi:10.1108/13612020610679268

Eppen, G. D., & Iyer, A. V. (1997). Backup agreements in fashion buying – The value of upstream flexibility. *Management Science*, *43*(11), 1469–1484. doi:10.1287/mnsc.43.11.1469

Forza, C., & Vinelli, A. (2000). Time compression in production and distribution within the textile-apparel chain. *Integrated Manufacturing Systems*, *11*(2), 138–146. doi:10.1108/09576060010314134

Geyskens, I., Stennkamp, J. E. M., & Kumar, N. (1999). A meta-analysis of satistaction in marketing channel relationships. *JMR, Journal of Marketing Research*, *36*(2), 223–238. doi:10.2307/3152095

Gripsrud, G. (2004). The marketing discipline and distribution research: Time to regain lost territory? In Håkansson, H., Harrison, D., & Waluszewski, A. (Eds.), *Rethinking marketing. Developing a new understanding of markets* (pp. 189–206). Chichester, UK: John Wiley & Sons.

Guercini, S. (2001). Relation between branding and growth of the firm in new quick fashion formulas. *Journal of Fashion Marketing and Management*, *5*(1), 69–79. doi:10.1108/EUM0000000007280

Guercini, S. (2008). Matching format strategy and sourcing strategy in clothing retail: A conceptual representation. *International Journal of Process Management and Benchmarking*, *2*(3), 185–196. doi:10.1504/IJPMB.2008.017959

Hammond, J. H. (1990). *Quick response in the apparel industry*. Cambridge, MA: Harvard Business School Case 9-690-038.

Hoover, W., Eloranta, E., Holmström, J., & Huttunen, K. (2001). *Managing the demand-supply chain*. New York, NY: Wiley.

Hunt, S. D., & Nevin, J. R. (1974). Power in channels of distribution: Sources and consequences. *JMR, Journal of Marketing Research*, *11*(May), 186–193. doi:10.2307/3150557

Hunter, N. A. (1990). *Quick response in apparel manufacturing*. Manchester, NH: The Textile Institute.

Irpet – Textrends. (1991). *Dove e come nasce il prodotto moda. Il caso del tessile-abbigliamento*. Milan, Italy: Franco Angeli.

Iyer, A. V., & Bergen, M. E. (1997). Quick response in manufacturer-retailer channels. *Management Science*, *43*(4), 559–570. doi:10.1287/mnsc.43.4.559

Jackson, T. (2001). The process of fashion trend development leading to season. In Hines, T., & Bruce, M. (Eds.), *Fashion marketing: Contemporary issues* (pp. 122–135). Oxford, England: Butterworth-Heinemann.

Jacobs, D. (2006). The promise of demand chain management in fashion. *Journal of Fashion Marketing and Management*, *10*(1), 84–96. doi:10.1108/13612020610651141

Jones, R. M. (2002). *The apparel industry*. Oxford, England: Blackwell Publishing.

Kerin, R. A., Hartley, S. W., & Rudelius, W. (2009). *Marketing*. Boston, MA: McGraw-Hill Irwin.

Lowson, R., King, R. E., & Hunter, R. E. (1999). *Quick response: Managing the supply chain to meet consumer demand*. Chichester, UK: John Wiley & Sons.

Mattila, H., King, R., & Ojala, N. (2002). Retail performance measures for seasonal fashion. *Journal of Fashion Marketing and Management*, *6*(4), 340–351. doi:10.1108/13612020210448637

Milgrom, P., & Roberts, J. (1988). Communication and inventories as substitutes in organizing production. *The Scandinavian Journal of Economics*, *90*(3), 275–289. doi:10.2307/3440309

Palmer, J. W., & Markus, M. L. (2000). The performance impact of quick response and strategic alignment in specialty retailing. *Information Systems Research*, *11*(3), 241–259. doi:10.1287/isre.11.3.241.12203

Park, H. (2004). US retailers' cooperation with manufacturer promotional support. *Journal of Fashion Marketing and Management*, *8*(4), 412–424. doi:10.1108/13612020410560009

Porter, M. E. (Ed.). (1985). *Competitive advantage*. New York, NY: Free Press.

Raven, B. H., & Kruglanski, A. W. (1970). Conflict and power. In Swingle, P. (Ed.), *The structure of conflict* (pp. 69–109). New York, NY: Academic Press.

Richardson, J. (1996). Vertical integration and rapid response in fashion apparel. *Organization Science*, *7*(4), 400–412. doi:10.1287/orsc.7.4.400

Sheridan, M., Moore, C., & Nobbs, K. (2006). Fast fashion requires fast marketing. The role of category management in fast fashion positioning. *Journal of Fashion Marketing and Management*, *10*(3), 301–315. doi:10.1108/13612020610679286

Stabell, C. B., & Fjeldstad, O. D. (1998). Configuring value for competitive advantage: On chains, shops, and networks. *Strategic Management Journal*, *19*(5), 413–437. doi:10.1002/(SICI)1097-0266(199805)19:5<413::AID-SMJ946>3.0.CO;2-C

Stockert, A. (2004). Fashion consumers. How to serve them if we don't know them. In Jacobs, D., & Sockert, A. (Eds.), *Fashion chachacha. Fashion chained and unchained: Chances and changes in the chain* (pp. 3–12). Amsterdam, The Netherlands: HvA Publicaties.

Tyler, D. J. (2008). *Carr & Latham's technology of clothing manufacture* (4th ed.). Oxford, England: Blackwell Publishing.

Uzzi, B. (1995). *Through the looking glass: The functional and malfunctional effects of embeddedness on economic action in New York's apparel industry*. Unpublished manuscript. Evanston, IL: Kellogg Graduate School of Management, Northwestern University.

Weiss, E. B. (1958). How much of a retailer is the manufacturer. *Advertising Age*, 29.

ENDNOTES

1 The study of the retail-apparel-textile system seems amenable to a functionalist approach: given its internal links to the supply channel, it can be analyzed as a whole (Bucklin, 1970). The sole fact that a number of different manufacturing agents are involved (in both textiles and apparel) does not by itself seem sufficient to compromise the application of some facets of functionalist analysis (after all, the distribution channel begins with manufacturers).

2 Good examples in this regard are the fashion-oriented product lines offered by Zara and H&M, or those of the Italian Liu Jo, Patrizia Pepe and Pinko.

3 The offer chain / distribution channel can be interpreted in both a "supply" and a "demand" perspective (Jacobs, 2006), as it presents aspects of both (Gipsrud, 2004, p. 202).

4 This may also represent an alternative response to the appeals of many retailers for ever lower prices for the manufacturing supplier's products (Guercini, 2008).

This work was previously published in Fashion Supply Chain Management: Industry and Business Analysis, edited by Tsan-Ming Choi, pp. 235-250, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 53
Production Information Systems Usability in Jordan

Emad Abu-Shanab
Yarmouk University, Jordan

Heyam Al-Tarawneh
Ministry of Education, Jordan

ABSTRACT

Enterprise systems are becoming more important as they support the efficiency and effectiveness of operations and reduce cost. In this chapter we explored the literature related to production information systems (PIS), enterprise systems, and other applications and their influence in an industrial zone in Jordan. Constructs from the Innovation Diffusion Theory were used, where results indicated that the adoption rate is acceptable and all variables have high means with respect to their evaluation by managers, but only two variable significantly predicted intention to use. In a second study that explored the status of IT usage in manufacturing firms using a different sample, results indicated that accounting information systems were widely used and distribution systems and manufacturing aiding systems were the least used. Other findings, conclusions and future work are stated at the end of the chapter.

INTRODUCTION

The industrial sector in Jordan is one of the main dimensions in the economic life of the country and is a major contributor in local production figures. Statistics indicate a contribution of 21.5% of the local production in the year 2006, which concluded 90% of the national size of exportation (Jordan Industrial Chamber Website, 2008).

DOI: 10.4018/978-1-4666-1945-6.ch053

Thus, the attention paid to this sector is one of the factors that lead to improving this sector's efficiency and productivity. Statistics also indicated that this sector employs 15% of The Jordanian workforce, which is a high percentage compared to other sectors (same source).

Research defined innovation as an intellectual performance that leads individuals to problem solving, or the intellectual effort that leads to non-repetitive or ordinary results (Trairy, 2008). Gnaim (2005) used another definition of innovation but in the higher education quality area, where he claimed that innovation is doing something good and not bad. Finally, Smadi (2001) stated that employee innovation in the manufacturing area in Jordan was more prevalent in small-size businesses than in larger ones, where he explored 870 employees and studied their inclination to adopt the Kaizen model in improving work environment.

One of the most important tools that help in improving this sector is information technology (IT), where its role in this sector ranges from supporting operations to a major role in automation and control of operations. The important role of information technology in manufacturing especially in reducing cost and pushing operations forward, and adding value to the manufacturing process, which satisfies firms' objectives. The importance of production information systems (PIS) is becoming vital to all manufacturing firms to gain and sustain competitive advantage. The adoption of PIS is becoming a priority to firms in this sector especially those who work in alliance with foreign (global) firms.

The word adoption is not a secret one, firms need to convince their employees and managers to effectively use and utilize information systems in a way that achieves firms' objectives. What makes this adoption more important, is that when we deal with complex systems (like EIS or PIS) users need more attention and involvement to guarantee their acceptance and thus the full benefit from such systems (Wang, Hsieh, Butler & Hsu, 2008). Research indicates the importance of users' adoption when dealing with new technology. Users are cautious to use new technology unless they believe it will bring to them some advantage, and would be easy to use. Also, they need to make sure it is compatible with systems they used before and many other factors that influence their perception of technology. This study reviewed the literature related to the technology acceptance area and adopted the innovation diffusion theory (IDT) proposed by Rogers in early 1983 (Rogers, 1995; Moore and Benbasat, 1991). This study reviewed also the literature in the area of production systems explored the factors predicting the rate of adoption of such systems through the reported opinions of managers in this sector. Finally, conclusions and future work are stated at the end.

BACKGROUND

Many theories tackled the adoption of new technology concepts and proposed a variety of theories and models that exceeded hundreds of propositions and variables. The argument in this domain emphasizes finding the suitable set of "factors" that can predict users' behavior with respect to using new technology. Most theories and research utilized the users' "intention to use" the technology as a surrogate to actual usage of the technology. Thus, many theories used "Intention to Use" (ITU) as the dependent variable, and proposed many predictors ranging from two variables to more than ten in some cases. The following section will explore the literature related to the IDT and other theories in the technology acceptance domain, then literature related to production information systems and enterprise resources planning systems usage.

Production Information Systems

Enterprise Information Systems (EIS) control all major business processes with a single software architecture in real time (Turban, Leidner, McLean & Wetherbe, 2008). Under the EIS category, the same authors list production information systems (PIS), Supply Chain Management Systems (SCM systems), Customer relationship Management Systems (CRM systems), Product lifecycle management systems (PLM systems), and Enterprise Resource Planning systems (ERP Systems). EIS account for 54% of licensing revenues, and expected to have the highest growth rates (expected to be \$55 billion in 2012) (McCrea, 2008). Our focus in this study is more towards PIS and ERP as they relate to manufacturing and industrial applications.

Production Information Systems (PIS) are defined as systems that work with production and operations information. PIS collect information from end terminals (like point of sale terminals - POS, shop floor machines, and operation and factory sensors) and store in transaction processing machines and then feed some management information systems (MIS) or other types of functional systems like: decision support systems (DSS), and enterprise resource planning systems (ERP) (Turban, Leidner, McLean & Wetherbe, 2008). The output of PIS is used to support managers in the process of decision making and to improve the managerial functions in the firm. The described architecture is not a unique one, as many disputes are related to the difference between functional systems and management information systems, as an example, PIS can get input data from transaction processing systems (TPS) and generate reports and output that can be used in the managerial decision making process. Some academicians describe all these systems as MIS, even enterprise resource planning systems (ERP).

Many definitions were reported in the literature for PIS like the definition used by Ciurana, Garcia-Romeu, Ferrer and Casadesus (2008) which indicates that PIS are systems that are related to transferring raw material into products with special specifications.

PIS are defined as data processing network systems (Hssain, Djeraba & Descotes-Genon, 1993, pp. 1). It seems that it is a simple definition but, the authors add further that it includes other components like the input data from process devices (sensors and terminals). On the other hand, the output is meaningful information used in decision-making. The authors report three methodological requirements related to design of PIS: managing the large volume of data and knowledge compounding the core of the network, coping with high complexity of operations, and providing reliable and available data and knowledge inputs to the PIS. Hsu and Rattner (1990) conclude that to meet these requirements, design of PIS should be a part of a global design approach within total-system architecture.

PIS aim at planning and scheduling and organizing for operations with related issue of work orders to shop floors and production (Ciurana, Garcia-Romeu, Ferrer & Casadesus, 2008). The importance of PIS comes from the role that it plays in facilitating the process of design and production of products, and forwarding it to customers (the distribution function). Research related to PIS focused on the applications related to the theory of planning and production control, where the main objective is reducing costs and risks, this is applied in more than one industry and type (Wang & Hu, 2008). On the other hand, research stressed the importance of providing the correct and timely information through the installation of sensors on the production line, or the use of web-based systems. Research emphasized the importance of synchronization of information systems to guarantee the required reduction of cost and the needed customer satisfaction (Mourtiz et al., 2008).

PIS can aid in the reduction of cost and facilitation (smoothing) of the flow of material in the manufacturing process. Also, the flow of information and material is becoming more vital to the production process when considering the global perspective or the supply chain management concept (SCM). A study that explored mobile communication technology importance in the integration process when transferring information and material between suppliers and customers, concluded that it is very important to utilize the benefits of PIS to reduce cost and gain integration between partners (Ende, Jaspers & Gerwin, 2008). In another study that explored the application of fuzzy logic concepts within a smart agent, solutions were provided for production problems utilizing previous solutions to previous problems (Lu & Sy, 2008). Authors concluded that it is useful in the industrial environment to provide decision makers with information residing in PIS to make accurate and timely decisions.

Finally, a group of Greek researchers studied the manufacturing strategies and the influence of IT on financial performance, where cluster analysis method with VACOR algorithm was used. Results indicated that the usage of information technology in industrial sector would have a significant effect on financial measures especially in organizations utilizing flexible manufacturing and with medium cost. The influence of using IT was higher in organizations that concentrate less on innovation and quality (Theodorou & Florou, 2008).

ERP Systems Research

The American Production and Inventory Control Society (APICS, 1998) defines ERP as *"an accounting-oriented computer information system that assists enterprises to define and plan on the resources required during the operation process of purchase, production, distribution and strategic planning to satisfy customers' orders"* (Lo, Tsai & Li, 2005, pp. 13). Deloitte Consulting (1999) refers to ERP systems as a packaged business solution that is designed to automate and integrate business processes, share common data and practices across the enterprise and provide access to information in a real time environment.

The previous definitions present a link between PIS and ERP systems, where the operation and production are closely linked through both types of systems. Thus, this study will visit literature related to ERP systems that might built a better picture of users' adoption of these systems.

ERP developers claim that their systems are claimed to increase firms' productivity, improve operational efficiency, improve communication and collaboration, grant better control on processes, enhance scalability and performance, leverage IT infrastructure, and reduce cost (SSA Report, 2006). ERP systems expand the processes of operations to reach external partners as well as internal firm constituents (Microsoft, 2003) and increase the level of automation and the value of business (SAP, 2006). Oracle proposed an enterprise manufacturing intelligence (EMI) tool that adds contextual dimension to shop floor data then synchronizes with ERP systems. Such addition provides real time intelligence to operations and move manufacturing and resource planning to another level (Smith, 2008).

When looking at research work, Carton and Adam (2008) concluded that ERP systems improved management decision making process. The authors did a case study in the pharmaceutical sector and concluded to the fact that implementing and utilizing ERP systems faces many barriers that prevent firms from benefiting from the rich information like: the lack of skills needed and the lack of access to data. Also they emphasized the incompatibility of transaction data with aggregate management queries.

Research work related to the perceptions of users towards ERP systems implementation explored different factors that make ERP systems more acceptable. Fan and Fang (2006) tested empirically the DeLone and McLean model. Their main findings were the significant relationship between system quality and system use; system quality and perceived usefulness; information quality and user satisfaction; perceived usefulness and system use; perceived usefulness and user satisfaction; system use and individual impact; user satisfaction and individual impact; and finally, individual impact and organizational impact. The major conclusion of the study was that user perceptions of IS success plays a critical role in ERP implementation. Similar to their study, Singla (2005) explored ERP performance and found that adopters consistently outperformed non-adopters across a wide variety of measures. The study used tangible and intangible benefits of ERP and a set of business performance measures.

Some research explored the resistance to change when applying and ERP systems. The study utilized an empirical test using a sample of Dunlop C&F employees (30 responses). Results indicated that firms need to work towards improving the user's attitude towards the ERP system. Again attitude is affected by users' self-efficacy and perceived usefulness. Other issues that can inflect failure into the implementation process might be employee education and vendors' support (Tsai & Hung, 2008). ERP systems integrate resources to enhance the overall performance and reduce costs (Lo, Tsai & Li, 2005). In a try to differentiate between ERP adopters' performance and non-adopters' performance, Hunton, Lippincout and Reck (2003) reported a decline in performance over time for non-adopters, while adopters enjoyed a steady performance.

We conclude that it is crucial to benefit from the capabilities of PIS, ERP and other IS in the industrial sector to gain competitive advantage in the market through the reduction of cost and the smoothing of material and information flow, and better utilization of resources.

Innovation Diffusion Theory

Rogers's diffusion of innovation (DOI) model appeared to be one of the most widely accepted models by researchers in identifying the "perceived" critical characteristics for innovation in information systems research and technology adoption research.

Rogers (1995) defines the "innovation diffusion process" as the spread of new ideas from its source of invention or creators to its ultimate users or adopters. Innovation is 'an idea, practice or object that is perceived as new by an individual or other unit of adoption (Rogers 1995, p.11). Diffusion is the 'process by which an innovation is communicated through certain channels over time among the members of a social system (Rogers 1995, p. 5).

Rogers classified people into categories according to their innovativeness; where individuals who are relatively earlier in adopting new ideas than other members of their social system are considered to be more innovative. The categories are the following arranged from adopters to non-adopters: innovators, early adopters, early majority, late majority, and laggards.

According to Rogers' theory, people start to adopt innovations slowly, and then the increased number of adoptions will continue until it reaches a peak, after that it diminishes leaving few individuals as non-adopters (laggards). Innovators are willing to try new ideas and venturesome. Early adopters are the opinion leaders in their community and carefully adopt new ideas. Individuals in the "early majority" category adopt new ideas before the average person and rarely seen as a leaders. On the other hand, individuals in the "late majority" category adopt an innovation only after a majority of people have tried it. At the end come the laggards who adopt the innovation only when it is a common and traditional aspect of their life or work. People in this category are suspicious of change, and resist such initiatives.

The Innovation Diffusion Theory (IDT) is a well accepted model in social sciences as it investigates the environment of adopting new innovations (the technology or information systems in this case) in organizations. The IDT was first proposed by Rogers (1983) and then modified by many researchers who modified the model and proposed different set of variables (Moore & Benbasat, 1991). Rogers (1983) proposed his theory as a model that includes the factors influencing the usage and adoption of innovation.

Rogers proposes five important factors that influence the adoption decision: the first is *relative advantage*: the degree to which the innovation perceived as being better than its precursor/supersedes. Second, *compatibility*: the degree to which an innovation is perceived as being consistent with existing system of values, past experiences and needs. The third is *complexity*: the degree to which an innovation perceived as being difficult to understand and use. Fourth, is *triability*: the degree to which an innovation can be tried and experimented with a limited basis. Finally, *observability*: the degree to which the beneficial results of use are observable and visible to users. Rogers addressed that innovation was more likely to succeed and be more readily adopted if the relative advantage as a consequence of its introduction was evident; if it was compatible with the adopter, its operations and its view of the world; if it was not 'too' complex; and if it was trialable and results can be observed prior to adoption.

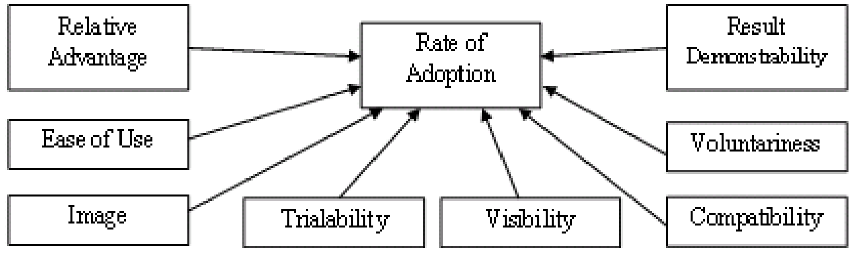
The theory measures three major areas: the adoption rate of technology related to time, where the theory is a suitable tool for measuring the diffusion of innovation in organizations (Brancheau & Wetherbe, 1990). Later, some researchers utilized this theory in studying the gap in technology accommodation (Fichman & Kemerer, 1999). Second, the work of Brancheau and Wetherbe (1990) included demographic factors related to the innovator characteristics that were explored in addition to the original factors proposed in the original theory. Brancheau and Wetherbe concluded that younger individuals were more receptive to technology adoption earlier in the process. Also, better educated individuals (subjects with higher degrees of education) were more open to interacting with the technology and thus were the ones with initiatives and opinion. Finally, the same authors explored the adoption of technology, where they anticipated that adopting a technology would go through four stages: awareness, persuasion, decision and implementation. The work of Brancheau & Wetherbe, coincides with the work of Agarwal (2000) with respect to the four adoption stages reported. Agarwal proclaimed that the IDT provides a better explanation of the adoption process and its interaction with time.

The work done by Moore and Benbasat (1991) is considered one of the important milestones in the life of this theory, where they used Rogers' work to develop a well validated instrument to measure the factors involved in predicting the rate of adoption. Moore and Benbasat built an instrument that explains the rate of adoption with reliable level of measurement. They added at the first stage of their study two major constructs and found them to be distinctive from the five proposed ones in Rogers' work. The first was *Image*; defined as "the degree to which use of an innovation is perceived to enhance one's image or status in one's social system". The second is *voluntariness of use*; defined as "the degree to which use of the innovation is perceived as being voluntary or of free will" (Moore & Benbasat, 1991, pp. 195).

The constructs' formulation took a multi-stage process and resulted in an eight construct model predicting the rate of adoption. Researchers focused their efforts on the time extension of the adoption process, where they used a longitudinal perspective instead of the snapshot view (cross-sectional snapshot of the data). The model utilized in their work is depicted in Figure 1.

Moore and Benbasat built their results on the responses of 540 employees from more than one organization, where the personal work-station was the technology under consideration. The major objective of the study was to build a highly reliable instrument to measure the set of

Figure 1. The IDT model proposed by Moore and Benbasat (1991)



factors used in the IDT. The study included four stages; the first was to review all the literature and the instruments available and used at that time. The second stage, researchers reviewed the items used for each variable in the instrument using a panel of experts. The third stage performed a pilot test on 20 subjects, followed by a test on 66 subjects to improve the instrument and increase its validity and reliability. Finally, the instrument was used on a large scale, where 34 items were the results of this study that measured the variables of the study.

In later studies, Agarwal and Parasad (1998) used part of the variables of the IDT like relative advantage, ease of use and compatibility, to study personal innovativeness influence on the rate of adoption of new technology. The results of the study indicated that personal innovativeness will be a significant moderator in the relationships influencing the rate of adoption of new technology. On the other hand, other studies that depicted the set of factors used in the IDT showed that this theory is stronger in predicting the rate of adoption than one of the widely used models in the area; the technology acceptance model-TAM (Davis, 1989). Results indicated that the IDT explained 45% of the variability in the rate of adoption, where the TAM explained only 32.7% of the variability of the same construct (Plouffe, Hulland & Vandenbosch, 2001).

Other studies that adopted some of the variables listed in the IDT and proposed by Moore and Benbasat (1991) went and tested the tendency to purchase on the Internet (Fitzgerald & Kiel, 2002). Using a snowball sampling method, 128 respondents were recruited to complete a survey that included measures from three domains: perceived attributes of use (included six constructs from the IDT and perceived risk), traditional normative beliefs (partner, family, friends, and near peers), and Internet normative beliefs (e-mail, discussion groups, virtual communities, and chat rooms). The mediators in the model were attitude, and motivation to comply with normative beliefs, and the dependent variable was future use intent. The results indicated that attitude was a strong factor explaining the intent to use for both adopters and non-adopters. Also, the major constructs explaining attitude were result demonstrability and risk for adopters, and risk for non adopters.

Mirchandani and Motwani (2001) performed a study related to the usage of e-commerce activities and focused on the relative advantage and compatibility, but added few constructs like: enthusiasm of top manger, compatibility with the company, relative advantage, and knowledge within the firm. Speier and Venkatesh (2002) proposed a comprehensive model to explore the perceptions of salespeople that affected their decision to reject a technology. The study analyzed the responses at two points of time utilizing 454 salespeople across two firms that implemented sales force automation tools. The model tested the interactions between individual characteristics, role perceptions, organization's characteristics, individual perceptions of the technology, professional state, person-fit technology, objective outcomes, and subjective outcomes. Relative advantage was the only significant factor from the set of individual perceptions of the technology that affected job-fit (at both time measures). Voluntariness had a low effect on the individual perception factors. Finally, the study reported a failure of the sales force automation technology based on the rejection of the salespeople as a result of negative job-related perceptions.

In an integration of multiple models in the technology acceptance area, Hardgrave, Davis, and Riemenschneider (2003) integrated the TAM/TAM2, IDT, and TPB/TRA to come up with a model that consisted of perceived usefulness, complexity, compatibility, social pressure and organizational mandate (voluntariness) as predictors of intention to follow a methodology. The results confirmed the effect of all factors except complexity (ease of use). The authors redefined the model to account for the indirect effects and proposed another model that linked complexity to usefulness, which yielded a significant result. Also, social influence and compatibility kept their significant relations with intention and usefulness.

Finally, in Saudi Arabia, Al-Gahtani (2003) tested a subset of the IDT constructs in a study aimed at computer technology adoption by Saudi workers in 136 organizations. The usable responses were 1190, and the dependent variable was adoption of computer technology. The results confirmed the five proposed constructs adapted from the IDT (relative advantage, complexity, compatibility, trialability, and observability).

JORDANIAN MANUFACTURING FIRMS ADOPTION OF PIS

This paper investigated the extent that manufacturing firms in Jordan are adopting IT as one of the innovation tools that help in the production and operation processes. PIS are considered, for the purpose of this research, the innovation under consideration. This research consists of two tests, where a group of surveys were distributed on factories mainly in the Northern region of the country (mainly in Al-Hasan Industrial Zone). The total collected surveys were 74 surveys that included a number of questions related to the information technology and applications content of the firm and some information about the organization itself and their information systems usability. On the other hand, another survey was distributed in Al-Hasan Industrial Zone and only to general managers, information managers and heads of departments. The second survey focused on managers' perceptions about adopting this type of systems and how well they accepted such technology. Total number of surveys collected was 91 usable surveys. The two studies were not addressing the same sample, and results of both studies are not closely related.

Jordanian PIS Usability

Table 1 shows the demographic details of the sample distributed, where 100 surveys were distributed and 74 were collected (response rate 74%). The survey aimed at collecting data related to managers' opinions on the usage and availability of PIS and other related systems. The main objective was to measure the adoption of PIS in the manufacturing sector in Jordan. When asking managers about using IT in their operations, 66 factories out of 74 reported that they used one or more of the systems listed in Table 2. Also, 8 factories only reported that they did not use any type of IT or systems in their operations. Table 2 lists different types of systems and the users among Jordanian factories.

Managers were asked about the relationship between supply chain management systems (SCM) and the employment of PIS, and how these systems are related to suppliers' capabilities also. Results indicated that 39 factories employ central

Table 1. Information related to the factories studied

Number of factories according to number of employees		Number of factories according to sales size (in 1000JD)		Number of factories according to number of computers	
Employees	**Factories**	**Sales**	**Factories**	**Computers**	**Factories**
1-50	31	Less than 100	10	Less than 10	36
51-100	7	100-1000	16	10-100	22
101-200	7	1000-10000	13	> 100	10
201-1000	20	> 10000	8	No Info.	6
>1000	9	No Info.	27	**Total**	**74**
Total	**74**	**Total**	**74**		

Table 2. Detailed type of information systems used in Jordanian factories

#	Item (question asked)	Number of factories
1	Do you use any type of Information Technology	66
2	Do you use accounting information systems	60
3	Do you use special sales systems	42
4	Do you use production information systems	45
5	Do you use inventory and warehousing systems	47
6	Do you use computer-aided design systems	23
7	Do you use human resource and salary systems	52
8	Do you use quality assurance/control systems	36
9	Do you use Distribution systems	25
10	Do you use procurement information systems	37
11	Do you use manufacturing aiding systems	25

systems, 13 factories employ distributed systems, but integrated together. Also, results indicated that 20 factories employed enterprise systems that are utilized in many functions and tasks. Finally, only 8 factories have extended systems that reach suppliers, distributors and customers (SCMS or ERP).

The second part of the survey included items related to the systems used in these factories. 47 managers indicated that they are interested in extending and using part or all of the systems listed in Table 2. Also, the distribution of the source and type of these systems were as follows: 32 factories used locally designed systems, and 33 factories used ready-made exported systems (off-shelf systems).

One of the objectives of this study was to see the relationships between variables like computer diffusion, sales and employee size. This was done through the correlations between those variables to test if any relationship exists between those. Also, to relate demographics with the main objective of this work, we estimated a new construct based on the count of the number of systems employed by the firm (example: manager of firm XYZ checked yes for using three systems (accounting information systems, HR systems and sales systems), so the total number of systems employed were three). This number (or set of data) was correlated against each of the variables mentioned. The correlations matrix is depicted in Table 3.

The results indicated significant correlations between the three variables and the total number of systems deployed. Also, it is shown that significant correlations existed between all three variables.

The Intentions of Managers to Adopt PIS

The second major objective of this study was to explore managers' intentions to adopt or continue using PIS systems utilizing Rogers' IDT model. In a separate study, the researcher employed a

Table 3. Correlations matrix of the demographics against the number of systems employed

	Number of employees	Sales size	Number of computers
Number of employees	1		
Sales size	0.551**	1	
Number of computers	0.636**	0.810**	1
Total number of systems	0.437**	0.394**	0.398**

** Correlation is significant at the 0.001 level survey that introduced a description of PIS and asked questions related to the different constructs of the model. The main source for the items used in the study was Moore and Benbasat (1991) work. The items were translated to Arabic language and tested using 10 experts for language. The nature of this exploratory study makes it convenient to use such method and the size of data allows for such test. Data used a seven point Likert scale, where 1 indicates a high disagreement to the statement and 7 indicates a high approval to the statement. The total number of surveys collected was 91 surveys from factories in Al-Hasan Industrial Zone (total number distributed = 100).

The survey included 3 items for measuring rate of adopting, 5 items for measuring relative advantage, 3 items for measuring compatibility, 3 items for measuring image, 2 items for measuring voluntariness, 2 items for measuring trialability, 2 items for measuring visibility, 4 items for measuring results demonstrability, and 4 items for measuring ease of use. Table 4 shows some descriptive statistics related to constructs.

On the other hand, correlations between all variable depicted in the IDT model were calculated and they are shown in Table 5. All correlations were significant except two and as shown in the matrix (with different levels of significance). Also, all variables were entered to calculate the regression coefficients between the rate of adoption and all variables. Results indicated that a significant correlation exists between the variables and the rate of adoption, where the coefficient of determination R^2 = 27.6%, with a p value less than 0.001 ($F_{8,82}$ = 5.285, p < 0,001). Results are shown in Table 6.

DISCUSSION OF RESULTS

This exploratory work tried to answer two major questions using multiple studies and methods. The first objective was to explore the extent to which PIS are used and adopted in manufacturing companies in Jordan. Through a descriptive survey, opinions were collected from managers of 74 factories in an industrial zone in Jordan. Results indicated that 66 factories (89%) used at least one system related to production and operations. The most popular systems used in Jordan were accounting information systems (60 factories, 81%), and the least used systems were computer aided design systems (23 factories, 31%). Results indicated also that PIS were used in 45 factories (61%), and inventory and warehousing systems were used in 47 factories (63.5%). The results indicated a fair adoption rate for such systems. Part of the reason for that is the influence of partnership with global firms and international organizations that outsource part of their production within Al-Hasan

Table 4. Descriptive statistics related to constructs in the IDT model

Variable	Number of Surveys	Min	Max	Mean	Standard Deviation
Rate of adoption	91	1	7	4.762	1.775
Relative advantage	91	2	7	5.777	1.123
Ease of use	91	1	7	5.409	1.334
Image	91	1	7	4.538	1.505
Compatibility	91	2	7	4.597	1.362
Result demonstrability	91	1	7	4.797	1.517
Visibility	91	1	7	4.637	1.540
Trialability	91	1	7	5.588	1.303
Voluntariness	91	1	7	4.654	1.615

Table 5. Correlation matrix showing the IDT variables

	RoA	RA	EoU	I	C	RD	V	T	V
Rate of adoption (RoA)	1								
Relative advantage (RA)	.271**	1							
Ease of use (EoU)	.030	.332**	1						
Image (I)	.386**	.397**	.481**	1					
Compatibility (C)	.351**	.269*	.419**	.637**	1				
Result demonstrability (RD)	.440**	.281**	.489**	.579**	.598**	1			
Visibility (V)	.310**	.256*	.367**	.465**	.578**	.495**	1		
Trialability (T)	.289**	.244*	.301**	.292**	.460**	.337**	.635**	1	
Voluntariness (V)	.207**	.156	.390**	.274**	.455**	.422**	.526**	.453**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Industrial Zone. It also seems that accounting and human resources systems were more popular as it is a major function for any firm regardless of their size of operations. This conclusion might be as a result of another test we did on the size of firms with respect to their sales and number of employees.

Using supply chain management systems (SCMS) was not that popular as only 8 firms used an extended system (11%), and the reason for that are the distinctiveness of the sample, as the sample in the first study came from an industrial zone, where international contracts are more common and a closed system from the local market is forced in this free zone. Such situation might be in a lesser need to an extended integrated system (SCMS). Finally, enterprise systems were not that popular as only 20 factories indicated using such integrated type of systems (27%)

When trying to explain the results of the correlations between the total number of systems (a measure of usability in this study) and the number of computers and employees and the total sales, it seems obvious that the size of the firm is a

Table 6. Coefficients table for the multiple regression test

Variable	Beta	Std Error	Std Beta	t	Sig
Constant	1.334	1.026		1.301	0.197
Relative advantage	0.246	0.158	0.156	1.556	0.124
Ease of use	**-0.506**	**0.148**	**-0.380**	**-3.411**	**0.001**
Image	0.269	0.156	0.228	1.720	0.089
Compatibility	0.016	0.178	0.013	0.092	0.927
Result demonstrability	**0.444**	**0.146**	**0.380**	**3.038**	**0.003**
Visibility	-0.010	0.156	-0.009	-0.067	0.947
Trialability	0.210	0.164	0.154	1.283	0.203
Voluntariness	0.041	0.125	0.037	0.329	0.743

Dependent variable: Rate of adoption, method: enter direct influencer (predictor) of the usability of IT. The larger firms will have higher numbers of employees and larger sales and thus they tend to utilize technology to improve operations and gain competitive advantage in the market. Also, it is logical to conclude that the firm size is directly correlated to complexity of operations and thus firms adopt IT to better control operations and improve flow of material and information. Finally, we can conclude that firms with higher sales will have larger tendency to invest in IT and thus buy more computers and adopt more types of systems.

The second objective was to explore managers' intention to adopt such systems. Results indicated a high intention to adopt PIS because of two main reasons: The first was the high mean values of predictors, which indicates the high perceptions of managers towards the adoption process. All means were above 4.5 out of 7, which indicate a high acceptance rates with respect to all variables used. The second reason for this conclusion is the high significant bivariate correlations with the rate of adoption, and this indicates that the method used and the large number of predictors was a limitation. The only variable with none significant correlation is ease of use and this supports the limitation of the method used. The highest correlation was between rate of adoption and results demonstrability (0.440**).

On the other hand, when summed together, the set of variables competed on the variance and only two variables showed significant prediction of rate of adoption. The two variables are: results demonstrability; the ability to see tangible results out of the system, and ease of use; where the complexity of the system is a huge obstacle to using it. The IDT model explained 27.6% of the variance in the rate of adoption. Results might have some limitations as the IDT have 8 predictors competing on the variance in the rate of adoption and this might limit the ability to explain the dependent variable well. The regression method used was to enter all variables forcefully and this might be the reason behind this surprising result. As this study is an exploratory one, we can conclude that a larger sample size and a thorough conceptual analysis of the predictors will lead to better utilization of variables and better and accurate results.

CONCLUSION

This paper aimed at exploring the status of using IT and specifically PIS in the area of production and manufacturing in Jordan. The study utilized two samples for two separate studies; the first was a sample of managers mainly related to IT in a group of factories in Al-Hasan Industrial Zone and other areas mainly in the Northern part of Jordan to explore the usage of PIS and other types of systems in the industrial area. The second study utilized another sample (after four months and from a different set of factories), from the same area and from the Northern part of the country to test the IDT using an instrument translated from Moore and Benbasat work (1991). Results indicated that systems like accounting information systems and HR and payroll systems were the mostly used among firms and distribution and manufacturing aiding systems were the least used among the sample used. The role of IS in production area was highly appreciated and a major conclusion is that the size of firm indicates the high computer usage and the diversity of systems used.

The second study resulted in high and significant indicators in predicting the adoption rate, and most of the constructs used in the IDT were significantly correlated to rate of adoption. But when regressing all indicators on rate of adoption, only two competed on the variance and yielded significant explanation of the variability of the dependent variable and they were results demonstrability and ease of use.

One of the limitations of this research, which makes its generalizability limited, is the usage of two separate samples. This research utilized two different samples, and to relate the real usage of PIS to the adoption rate, the same sample would have been used. Still the inferred results of this work are valid, but researchers are encouraged to use one sample and extend the size to improve the statistical generalizability. The second limitation of this study is the instrument used; this study used a translated instrument from the original one used in Moore and Benbasat (1991) in English, and thus researchers are encouraged to use the instrument in Arabic to improve the language and improve content and face validity of the instrument. Finally, research related to PIS and the factors influencing the adoption of such systems is not highly popular, which resulted in high competition between variables. The IDT needs a larger sample or dropping some of the variables based on conceptual bases. When exploring systems like ERP or PIS systems, as they are considered complicated and comprehensive systems, one needs to keep relative advantage and ease of use for sure, but further exploration needs to be done to try to deduct the scale size and improve predictability of the model.

FUTURE RESEARCH DIRECTIONS

This research is needed in this area and considered a first step in validating the instrument and testing factors influencing the rate of adoption. It is highly important to continue such research using longitudinal settings to explore the adoption and check the validity of results.

Future research is needed to validate the instrument and apply it to more settings and environments. Another direction that is needed is the multi-stage process applied by Moore and Benbasat (1991), where the adoption rate is investigated with time and also, a better conceptual perspective is reached through the reduction of variable. One idea is to compare other models predictability with the IDT like the Technology Acceptance Model (TAM), the theory of Reasoned Action (TRA), the Theory of Planned Behavior (TPB) and its extension the Decomposed Theory of Planned Behavior (DTPB).

As we now know the situation of PIS usability in industrial zones, would that knowledge facilitate better research in other environments like local industrial areas and other major factories in Jordan? Also, would it result in a different conclusion if we explored other types of systems? Finally, results indicated a weakness in utilizing computer aided design systems, future research can explore the reasons behind such phenomenon and would that be related to the industrial development of the sector in general or because of this global partnership with local factories specifically in the industrial zone.

REFERENCES

Agarwal, R. (2000). Individual acceptance of information technologies. In Zmud, R. (Ed.), *Framing the domains of IT management* (pp. 85–104). Cincinnati, OH: Pinnaflex Education Resources, Inc.

Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*, *9*(2), 204–215. doi:10.1287/isre.9.2.204

Brancheau, J. C., & Wetherbe, J. C. (1990). The adoption of spreadsheet software: testing innovation diffusion theory in the context of end-user computing. *Information Systems Research*, *1*(2), 115–143. doi:10.1287/isre.1.2.115

Carton, F., & Adam, F. (2008). ERP and Functional Fit: How Integrated Systems Fail to Provide Improved Control. *The Electronic Journal Information Systems Evaluation, 11*(2), 51 – 60. Retrieved from http://www.ejise.com

Ciurana, J., Garcia-Romeu, M., Ferrer, I., & Casadesus, M. (2008). A Model for Integrating Process Planning and Production Planning and Control in Machining Processes. *Robotics and Computer-integrated Manufacturing*, *24*, 532–544. doi:10.1016/j.rcim.2007.07.013

Davis, F. D. (1989). Perceived usefulness perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, *13*(3), 319–340. doi:10.2307/249008

Deloitte Consulting. (1999). *ERPs Second Wave* [Report]. Deloitte Consulting. Retrieved from http://www.deloitte.com

DeLone, W., & McLean, E. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, *3*(1), 60–95. doi:10.1287/isre.3.1.60

Department of Statistics. Jordan. (2008). *Statistics related to the Jordanian Industrial Sector*. Retrieved from http://www.dos.gov.jo/dos_home_a/gpd.htm

Ende, J., Jaspers, F., & Gerwin, D. (2008). *Involvement of system firms in development of complementary products: The influence of novelty*. Technovation.

Fan, J., & Fang, K. (2006). ERP Implementation and Information Systems Success: A Test of DeLone and McLean's Model. In *PICMET2006 Conference proceedings,* July 2006,Turkey (pp. 9-13).

Fichman, R. G., & Kemerer, C. F. (1999). The illusory diffusion of innovation: an examination of the assimilation gaps. *Information Systems Research*, *10*(3), 255–275. doi:10.1287/isre.10.3.255

Fitzgerald, L., & Kiel, G. (2001). *Applying a consumer acceptance of technology model to examine adoption of online purchasing*. Retrieved February 2004, from http://130.195.95.71:8081/WWW/ANZMAC2001/anzmac/AUTHORS/pdfs/Fitzgerald1

Gnaim, K. (2005). Innovation is one of Quality aspects. *Quality in Higher Education*, *1*(2).

Gupta, S., & Keswani, B. (2008). *Exploring the Factors That Influence User Resistance to the Implementation of ERP*. Hyderabad, India: The ICFAI University Press.

Hardgrave, B. C., Davis, F. D., & Riemenschneider, C. K. (2003). Investigating determinants of software developers to follow methodologies. *Journal of Management Information Systems*, *20*(1), 123–151.

Hssain, A., Djeraba, C., & Descotes-Genon, B. (1993). Production Information Systems Design. In *Proceedings of Int Conference on Industrial Engineering and Production Management (IEPM-33),* Mons, Belgium, June 1993.

Hsu, C., & Rattner, L. (1990). Information Modeling for Computerized Manufacturing. *IEEE Transactions on Systems, 20*(4).

Hunton, J., Lippincott, B., & Reck, J. (2003). Enterprise Resource Planning Systems: Comparing Firm Performance of Adopters and Nonadopters. *Accounting Information Systems*, *4*, 165–184. doi:10.1016/S1467-0895(03)00008-3

Jordan Industrial Cities. (2008). *Statistics from the website of the JIC.* Retrieved from http://www.jci.org.jo

Lo, C., Tsai, C., & Li, R. (2005, January). A Case Study of ERP Implementation for Opto-Electronics Industry. *International Journal of The Computer*. *The Internet and Management*, *13*(1), 13–30.

Lu, K., & Sy, C. (2008). *A real-time decision-making of maintenance using fuzzy agent*. Expert Systems with Applications.

McCrea, B. (2008). ERP: Gaining Momentum. *Logistic Management,* November 2008, pp. 44-46.

Microsoft. (2003). *Microsoft Business Solutions*. Retrieved from http://www.microsoft.com/business solutions

Mirchandani, D. A., & Motwani, J. (2001). Understanding small business electronic commerce adoption: an empirical analysis. *Journal of Computer Information Systems*, *41*(3), 70–73.

Moore, G., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2*(3), 192–222. doi:10.1287/isre.2.3.192

Mourtzis, D., Papakostas, N., Makris, S., Xanthakis, V., & Chryssolouris. (2008). Supply chain modeling and control for producing highly customized products. *Manufacturing Technology Journal*.

Plouffe, C., Hulland, J., & Vandenbosch, M. (2001). Research report: richness versus parsimony in modeling technology adoption decisions-understanding merchant adoption of a smart card-based payment system. *Information Systems Research*, *12*(2), 208–222. doi:10.1287/isre.12.2.208.9697

Report, S. S. A. (2006). *SSA ERP on SOA Platform*. SSA Global and IBM. Retrieved from http://www.ssaglobal.com

Rogers, E. M. (1983). *The diffusion of innovations*. New York: Free Press.

Rogers, E. M. (1995). *The Diffusion of Innovation* (4th ed.). New York: Free Press.

SAP. (2006). *SAP Customer Success Story*. Retrieved from http://www.sap.com

Singla, A. (2005). Impact of ERP Systems on Small and Mid Sized Public Sector Enterprises. *Journal of Theoretical and Applied Information Technology,* 119-131.

Smadi, S. (2001). Employees' Attitudes Towards the Implementation of the Japanese Model Kaisen for Performance Improvement and Meeting Competitive Challenges in The Third Millennium: The Jordanian Private Industrial Sector. *Abhath Al-Yarmouk,* 313-335.

Smith, F. O. (2008 May). Oracle Says It Will Leapfrog Competitors in Manufacturing Intelligence. *Manufacturing Business Technology,* 26-29.

Speier, C., & Venkatesh, V. (2002). The hidden minefields in the adoption of sales force automation technologies. *Journal of Marketing*, *65*, 98–111. doi:10.1509/jmkg.66.3.98.18510

Theodorou, P., & Giannoula, F. (2008). Manufacturing strategies and financial performance- The effect of advanced information technology: CAD/CAM systems. *The International Journal of Management Science*, *36*, 107–121.

Trari, A. (2008). مكتبة جامعة اليرموك. Retrieved from http://library.yu.edu.jo/

Tsai, W., & Hung, S. (2008). E-Commerce Implementation: An Empirical Study of the Performance of Enterprise Resource Planning Systems Using the Organizational Learning Model. *International Journal Of Management, 25*(2).

Turban, E., Leidner, D., McLean, E., & Wetherbe, J. (2008). *Information Technology for Management* (6th ed.). Hoboken, NJ: John Wiley.

Wang, T., & Hu, J. (2008). An Inventory control systems for product with optional components under service level and budget constraints. *European Journal of Operational Research*, *189*, 41–58. doi:10.1016/j.ejor.2007.05.025

Wang, W., Hsieh, J., Butler, J., & Hsu, S. (2008). Innovative Complex Information Technologies: A Theoretical Model And Empirical Examination. *Journal of Computer Information Systems*, (Fall): 27–36.

This work was previously published in Enterprise Information Systems Design, Implementation and Management: Organizational Applications, edited by Maria Manuela Cruz-Cunha and Joao Varajao, pp. 270-286, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 54
Research into the Path Evolution of Manufacturing in the Transitional Period in Mainland China

Tao Chen
SanJiang University, China, Nanjing Normal University, China, & Harbin Institute of Technology, China

Li Kang
SanJiang University, China, & Nanjing Normal University, China

Zhengfeng Ma
Nanjing Normal University, China

Zhiming Zhu
Hohai University, China

ABSTRACT

Manufacturing transition is an important part of industrial upgrading. At present, Chinese scholars study the problem of manufacturing chiefly from two perspectives: The first is to discuss the status quo of Chinese manufacturing from the perspective of industrial competitiveness, with countermeasures put forward against manufacturing upgrading. The second is to directly discuss the upgrading of manufacturing from the perspective of global value chain, with the following proposal put forward: Chinese manufacturing upgrading should stretch from the low end to both ends of value chain. In addition, a discussion is also made to the role of producer services in promoting manufacturing, and the role of governmental regulations in upgrading manufacturing. Although these two perspectives are rational, they have some defects: Both of them are based on the hypothesis that the institutional environment in which manufacturing lies is stationary, and manufacturing is considered and measured with systems as exogenous variables; so the impact of institutional environment on manufacturing upgrading is overlooked. Based on reviewing previous literature, this chapter analyzes and discusses the path evolution of manufacturing in the transitional period in mainland China.

DOI: 10.4018/978-1-4666-1945-6.ch054

1 INTRODUCTION

Industry, especially manufacturing, is the foundation and pillar of national economy. To most developed and developing countries, the leading role and fundamental function of manufacturing cannot be replaced by those of any other industrial sector. Since reform and opening-up began over 30 years ago, especially since China joined WTO in 2001, Chinese economy has been developing very rapidly, along with the continuous rise of its economic aggregate. This is closely linked with the swift development of manufacturing. So to speak, the development of manufacturing supports the "ridge" of Chinese economy. The same trend has also occurred to other countries. The governments of many Western countries have again brought forward their plan of "reindustrialization", i.e. paying attention to the important contribution of manufacturing to economic growth. Therefore, China cannot develop its economy without the development of manufacturing. Instead, we must pay attention to the role and function of manufacturing.

At present, Chinese manufacturing develops very rapidly. In the past ten years, both the volume and value of production of Chinese industry have always been growing rapidly. If calculated as per a constant price, the annual average growth of the total production value of Chinese manufacturing from 1995 to 2003 was 14.53%. By 2003, the total production value of manufacturing had reached about 12.27 trillion yuan. According to the calculation of UN Statistics Division and Industrial Development Organization, the annual average growth rate of Chinese manufacturing from 1998 to 2003 reached as high as 9.4%, while the same figure for developing countries in the same period was only 4.4.%. The value of the total export volume of Chinese manufactured goods divided by the number of employees in manufacturing rose from 1763.92 US dollars in 1995 to 9570.09 US dollars in 2004; the annual average growth rate from 1998 to 2001 was 18.30%. Since China joined WTO, the growth rate of export has been increasing more rapidly. The annual average growth rate from 2002 to 2004 reached 22.50% (Jin,et al., 2007). In 2008, the number of manufacturing enterprises in China was 396950, with an added value of 44135.836 billion yuan, total assets of 32340.308 billion yuan, and 77315.7 thousand employees. The number of manufacturing enterprises, total production value, total assets and number of employees in 2008 grew by 174.90%, 497.13%, 243.69% and 67.37% respectively over 2000. Among them, the total production value had the biggest growth rate: nearly 500% (Lin, 2010).

However, behind the rapid development of Chinese manufacturing exists a series of problems yet to be solved in an effective way.

First, the per capita added value of Chinese manufacturing is far lower than the world average. If calculated as per the constant price in 2000, the per capita added value of Chinese manufacturing in 2006 was 610 US dollars, lower than that of the developing regions in West Asia and Europe, Latin America and the Caribbean, merely equivalent to 13.5% of that of industrialized countries (Please refer to Table 1) (Li, et al., 2009).

Second, the regional distribution and industrial structure of Chinese manufacturing are seriously imbalanced. From the perspective of regional distribution, a huge gap exists in the distribution of manufacturing among eastern, central and western China. The total production value, added value, the total assets and number of employees of the manufacturing in eastern China are 73.68%, 68.85%, 70.14% and 72.50% respectively; the same figures in central China are 16.23%, 18.86%, 16.91% and 16.75% respectively; the same figures in western China are 10.10%, 12.29%, 12.95% and 10.75% respectively. From the perspective of industrial structure, the distribution of manufacturing is also imbalanced. Nearly 66% of the added value of manufacturing in 2008 was distributed in ten industrial sectors including ferrous metal smeltering, communication equipment and computer (Lin,

Table 1. International comparison of the per capital added value of manufacturing (the constant price of 2000)

Year Country group	1991	1994	1995	1998	2000	2006
Industrialized countries other than CIS	3573	3614	3730	3996	4291	4509
African countries located to the south of the Sahara	29	26	26	27	27	30
East Asia and South Asia	128	151	164	166	199	267
Latin America and the Caribbean	664	694	680	736	739	792
West Asia and Europe	431	448	460	518	535	664
CIS countries	422	227	216	198	239	369
North Africa	164	162	167	183	195	207
China	167	233	254	313	366	610

Note: The data of China include those of Taiwan Province and Hong Kong Special Administrative Region, but exclude those of Macao Special Administrative Region.

Data source: UN Industrial Development Organization (UNIDO)

2010). This imbalance will lead to the huge waste of human, material and financial resources in eastern, central and western China, especially in central and western China, and the ineffective utilization of resources, thus setting back industrial development and economic progress.

Third, the added value of Chinese manufacturing makes up a relatively high proportion in China's GDP. From 2003 to 2007, the proportion of the added value of Chinese manufacturing in China's GDP did not change a lot, always remaining between 34% and 40%. This proportion was not only far higher than that of such developed countries as USA, Japan, Germany, UK and France, but also far higher than such developing countries as Brazil, India and Mexico. Although some research indicates that China is now in the mature period of new industrialization (Li Gang et al, 2009), but this relatively high index not only reflects the characteristics of industrial structure of China in the middle stage of industrialization, but also indicates that there possibly exists a problem of disproportion to some extent in Chinese economic structure (Jin, et al., 2007).

We should not only develop our economy, but also upgrade our manufacturing. We can solve the problems existing now in Chinese manufacturing only by transforming manufacturing from a low end to a high end and from resource wastage and environmental pollution to resource conservation and environmental protection by applying a series of practical measures. Therefore, so to speak, the sustainable, rapid and healthy development of Chinese economy in the future will be closely linked with the upgrading of manufacturing.

2 THE INTERNATIONAL COMPETITIVENESS OF CHINESE MANUFACTURING

Of all the fundamental theories of industrial international competitiveness having been so far put forward, the most influential are the theory of comparable advantage (Ricardo,1817) and the theory of competitive advantage (Porter,1990; Liu, et al., 2006). In Paul Krugman's International Economics (the most extensively-distributed and authoritative textbook in this field in the world), "comparative advantage" is defined as: "If the opportunity cost for producing a certain product in a country is lower than in other countries, then this country has relative advantage in producing this product". Therefore, the fundamental principle concerning comparative advantage and international trade is: "If a country exports commodities in which it has comparative advantage to another country, then both these two countries can benefit from the trade between them". In contrast, according to the theory of the competitive advantage of nations put forward by Professor Porter, the competition among countries in market economy in fact goes on among enterprises, and enterprises with international competitive advantage are concentrated in only a limited number of industrial sectors. Therefore, industrial sectors should be used as the basic units for studying national competitive advantage. The competitive advantage of enterprises not only comes from themselves, but also originates from the microeconomic foundations on which their development relies---a diamond system consisting of factor conditions, demand conditions, the competitive background of corporate strategy and structure and relevant supporting industrial sectors. Therefore, the attention to national competitive advantage should be focused on the cultivation of these microeconomic foundations (Liu, et al., 2006).

As regards to the index-based appraisal of industrial competitiveness, the industrial competitiveness of a country can be appraised by using more than one index. As indicated by relevant researches, we can derive somewhat different judgments if we appraise the competitiveness of Chinese manufacturing and its subsidiaries by using different indexes. Some indexes reveal that the international competitiveness of a certain industrial sector of China is elevated, while some other indexes reveal that the international competitiveness of this industrial sector of China is declining. As a matter of fact, this phenomenon often occurs when we observe a complicated thing from different perspectives, or the same complicated thing manifests itself in different ways in different aspects. Jin,et al.(2007) combine many indexes into one to express the trend of the change of the industrial competitiveness of China in a comprehensive way. Through due research, they worked out the composite index of the international competitiveness of Chinese manufacturing (Please refer to Table 2).

By comparing Chinese manufacturing with American and Japanese manufacturing, they formed such an opinion: The competitiveness of Chinese manufacturing has always been rising continuously. From the perspective of its subsidiaries, the competitiveness of Chinese manufacturing in "other products" in the above table is the strongest. Office products and telecommunication equipment, whose competitiveness has exceeded

Table 2. The composite index of the international competitiveness of Chinese manufacturing

	The index of comparative advantage	The index of competitive advantage	The composite index of the international competitiveness
Manufacturing	102.8	132.1	117.5
1.Iron and steel	210.0	172.5	191.3
2. Chemical finished products and relevant products	88.2	115.7	102.0
3. Other semi-finished products	98.1	130.5	114.3
4. Mechanical and transportation equipment	119.8	164.1	142.0
Including: Office products and electronic products	134.9	190.9	162.9
5.Textiles	93.5	132.8	113.2
6. Apparel	83.2	113.5	98.4
7.Other products	89.2	108.1	98.7

that of textiles, are called "the second most competitive industrial sector of manufacturing". Transportation equipment (especially automobile), chemical products, integrated circuits and components and power machinery and equipment have the weakest competitiveness in China, with their competitiveness still shrinking continuously. Compared with EU, China has the most obvious disadvantage in chemical products (especially medicines), transportation equipment (especially automobiles) and power machinery and equipment. In contrast, office and telecommunication equipment, other products (especially personal and household goods), apparels and textiles, enjoy a relatively high competitiveness advantage. At last they derived such a conclusion: The present elevation of the international competition of Chinese industrial enterprises is not determined by industrial enterprises to a very large extent; instead, this elevation depends on the development of Chinese finance and mass media. In our opinion, there are two important conditions for guaranteeing the enhancement of the international competitiveness of Chinese manufacturing: One is that Chinese manufacturing enterprises must secure a financial support from Chinese financial enterprises all over the world; the second is that China must have some media (including newspaper, TV and radio) influential in the mainstream crowds in foreign countries (especially European countries and the USA). Therefore, while assisting Chinese industrial enterprises in "going abroad", Chinese government must consider how to assist Chinese financial institutions (especially banks) and media (especially newspaper) in "going abroad".

In addition, Chen et al. (2009) studied the international competitiveness of Chinese and American manufacturing through empirical analysis on the basis of the hierarchy-based opinion of industrial competitiveness. In their opinion, industrial competitiveness consists of four hierarchies, which are in turn (from bottom to top): the source of competitiveness—industrial environment; the essence of competitiveness—productivity; the performance of competitiveness—market share; the result of competitiveness—industrial profitability. These four hierarchies are interlinked and cycled logically. The ultimate goal of industrial competitiveness is to generate profit; to generate profit, we should first prove that we have a stronger industrial competitiveness than other countries in trade; the foundation for enhancing trade competitiveness is to enhance the productivity of this manufacturing sector; to enhance the productivity, we should invest in the construction of soft and hard environments including technical innovation, advanced equipment and education & training; the capital resources for investment in environments rely on industrial profit in turn. Only by investing in environments (the first hierarchy) with profits (the highest hierarchy of competitiveness), can we enter the new-round cycle of productivity and market share, and can we acquire continuously-reinforced market competitiveness. On this basis, they analyzed the international competitiveness level of 30 types of manufacturing in China, and derived an inconsistent conclusion: Measuring industrial competitiveness with the index of profitability and the index of productivity has a very high goodness of fit, which proves that the industries with a higher productivity will also have a higher profitability in Chinese domestic market. However, the result of sequencing based on the index of profitability and the index of productivity is quite different from the result of sequencing based on the index of market share: Gamma coefficient is negative, which proves that the industrial sectors with a higher profitability and a higher productivity do not necessarily have a bigger global market share, or things may happen the other way round. This proves that the first hierarchy and explanatory variable of industrial competitiveness---the factor of industrial environments, exerts a very great influence upon the status of competitiveness; industrial structure, industrial policy, trade policy, system environment, etc. can bring about some difference among the three competitiveness indexes. Compared with Chinese manufacturing, not only are the indexes of the hierarchy of profitability and the hierarchy of productivity in American manufacturing interrelated, but also the competitiveness indexes of the hierarchy of market share and the hierarchy of productivity are somewhat interrelated; namely, the sequencing results of these three competitiveness indexes are not quite different. This proves that American manufacturing has a more mature development environment than Chinese manufacturing, so production efficiency can be smoothly transformed into market share and profit in the USA. From this we can see the importance of environments, namely, the important influence of such factors as systems upon the upgrading of manufacturing.

3 THE UPGRADING OF MANUFACTURING FROM THE PERSPECTIVE OF GLOBAL VALUE CHAIN

To discuss the upgrading of manufacturing from the perspective of global value chain is the mainstream of the contemporary research into industrial upgrading. As the internationalization of contemporary economy becomes more and more intensified, features that are different from those of the previous tide of globalization have occurred to production, trade, investment, turnover, means of organization, etc. As known to all, division of labor is the source of economic growth, while industrial transfer is an important means to realize spatial division of labor. As international division of labor is intensified from among industries to among different products in each industrial sector, and then to different working procedures of each product, international industrial transfer is also evolved from spatial transition among different industries to that among different products, and then to that among different working procedures of each product. The intensification of division of labor and industrial transfer are the key contemporary features of this globalization different from those of the previous rounds of globalization. On the other hand, global value chain is the leading force for boosting the intensification of division of labor and coordinating industrial transfer. This leading force has not only changed the microscopic foundation of globalization, but also produced a revolutionary effect upon competition models and development strategies.

As regards to the theory of global value chain, an American scholar (Michael E Porter,1985) put forward the theory of Value Chain from the perspective of the theory of enterprise competition strategies. This theory is a theoretical framework for analyzing enterprise activities under the condition of international competition. Its core opinion is: The value created by enterprises in fact originates from some specific value activities on a relevant value chain. Grasping “strategic link” is a key factor for controlling the entire value chain and relevant industrial sector. In 1990s, a scholar(Gereffi, Korzeniewicz,1994,1999)put forward the theoretical framework of Global Commodity Chain. This theory directly related the value-added chain with global industrial organizations, and therefrom made a comparative research into the commodity chain driven by producers and purchasers. To eliminate the limitation of the word “commodity”, and highlight the importance of the creation of the relative value of enterprises and the acquisition of value on the chain, Gereffi and numerous researchers in this field reached an agreement at the beginning of the 21st Century to replace global commodity chain with a term “Global Value Chain (GVC)”. The classification of global value chains is basically based on the dichotomy under the framework of global commodity chain, namely, the producer-driven value chain and purchaser-driven value chain put forward by Gereffi. The producer-driven value chain means boosting market demand through producers’ investment, thus forming a system of vertical division of labor of local production supply chain. Under this value chain, investors include not only transnational companies blessed with technical advantages and seeking for market expansion, but also national governments trying hard to boost local economic development and set up an independent industrial system. The purchaser-driven value chain means that large purchaser organizations blessed with strong brand advantages or sales channels coordinate and control production, design and marketing activities that aim at the target market. This value chain is featured by labor intensiveness and represented by consumer goods (e.g. apparel, shoe, toy, consumer electronics, etc.).

Another important content of the theory of global value chain is about the governance model of value chain. So far, no uniform conclusion has been reached in the academic circle on the classification of the governance models of global value chain. As regards to the governance models of global value chain, Humphrey, Schmitz(2002) made a distinction among the following four governance models on the basis of the difference of organization coordination and power distribution: market-based model; model based on equilibrium network; capture-based model; hierarchy-based model. The enterprises in developing countries generally enter the assembly and manufacturing link of capture-based GVC as subcontractors by relying on cheap labor force. The high competition and low income brought by low-entry barriers make subcontracting enterprises face a huge pressure of upgrading. Through an empirical analysis into the textile and apparel sector of industry in the world, Gereffi(1999)worked out the sequential upgrading model under GVC, namely technical upgrading—product upgrading---functional upgrading—chain's upgrading, and optimistically held the opinion that the subcontracting enterprises in developing counties can smoothly realize this kind of sequential upgrading by joining GVC and accepting the supports from leading enterprises in developed countries in such aspects as technical diffusion, employee training and equipment introduction. With this sequential upgrading, the performance of subcontracting enterprises, namely the quantity of value created and acquired by them, also increases gradually. Gereffi's analysis has two problems: First, as indicated by a lot of practice of developing countries, the above-mentioned model of sequential upgrading cannot be realized automatically (Humphrcy,Schmitz,2002). In addition, upgrading will change the contrast of powers and the structure of income distribution in GVC. Therefore, the upgrading of subcontracting enterprises is subject to suppression from leading enterprises. The size of upgrading barrier depends on the governance model of GVC. Next, upgrading does not necessarily mean the enhancement of the performance of subcontracting enterprises. In the capture-based GVC, the upgrading of subcontracting enterprises is a kind of passive upgrading aiming to obey the global strategies of leading enterprises. By continuously searching for and supporting new subcontractors and intensifying competition, leading enterprises capture the newly-added value created by the upgrading of subcontracting enterprises. In contrast, in the market-based and network-based governance with more equivalent powers, enterprise upgrading is a kind of active upgrading adapting to competition and seeking for profit, and can accordingly acquire the benefits brought by upgrading. According to the research of Liu (2007), and in connection with the practice of developing countries, Zhuo(2009) worked out four matching models of management, upgrading and enterprise performance: (1) Market-based management—(independent and slow sequential upgrading)—(slow enhancement of performance). Under this governance model, the transaction target is mature standardized products; the division of labor and transaction among enterprises are based on market contracts marked by "at arm's length"; therefore, enterprise upgrading is a kind of endogenous and independent upgrading based on competency. This upgrading is free from the control and hindrance of other enterprises, but it is relatively slow, so the enhancement of enterprise performance is also progressive. (2) Governance based on equilibrium network ---(independent and fast sequential upgrading)—(fast enhancement of performance). Governance based on equilibrium network is a coordination method of division of labor and transaction based on the mutual supplementation of competency, the sharing of knowledge and technology and relative equality; it does not involve a relationship between controlling and being controlled; therefore, enterprises can independently carry out upgrading in various forms. In this kind of network, the partial innovation required by high-degree division of labor greatly reduces the investment required by upgrading and mitigates investment risks, while the sharing of competency, technology and knowledge accelerates the realization of upgrading and the enhancement of performance. (3) Capture-based management---(rapid but passive technical upgrading and product upgrading)—(it is difficult to enhance the performance, or even the performance declines). In capture-based GVC, in order to guarantee the diversity of products, the timeliness of supply and the reliability of product quality, transnational companies of developed countries have to assist subcontracting enterprises in accelerating technical upgrading and product upgrading. In addition, by such means as patent pool, strategic isolation, brand reinforcement and retail market merger, they endeavor to raise the entry barriers for links with a high additional value including design, R & D and marketing, and slow down the process of the functional upgrading and chain's upgrading of subcontracting enterprises, so as to prevent their core competency and income from being eroded. Under such circumstances, it is generally difficult to realize the sequential upgrading of subcontracting enterprises, and the income created by the rapid technical and product upgrading is also captured by transnational companies as a result of the intensification of the competition in the assembly manufacturing link. (4) Hierarchy-based management--- (rapid but passive technical upgrading, slow product upgrading)—(it is difficult to enhance the performance greatly). Hierarchy-based governance model is a method of production organization and coordination. Under this model, in order to reduce their cost and occupy the market, transnational companies of developed countries establish joint ventures through FDI in other countries, and control and operate enterprises by relying on such core competencies as ownership and R & D design. In the hierarchy-based GVC, as joint ventures can directly obtain the technology, brands, capitals and equipment of transnational companies, they can realize technical upgrading rapidly, and make their products meet the uniform global quality standards of transnational companies. Under this model, product upgrading is relatively slow because of its dependence on the market development and competition status of the countries to which the investment is oriented. Moreover, functional upgrading and chain's upgrading are strictly controlled by transnational companies. On the other hand, it is difficult to greatly enhance the performance of joint ventures, because transnational companies try to squeeze the profit margin of joint ventures by collecting a high fee of technical transfer, key parts and components, and brand licensing.

With the Yangtze River Delta as its research subject, Liu, et al.(2009) analyzed the disadvantages of Chinese manufacturing upgrading under GVC. In their opinion, merging into GVC has weakened the relationship among industrial departments in different regions in China, and exerted an unfavorable influence upon the integrated development of regional substantial economy. The technology transfer and technology spillover embedded in outsourcing and subcontracting activities have an obvious "breakpoint" and "isolation" effect upon the industrial development of developing countries. The imbalance and asymmetry of GVC income distribution lead to the difficulty in fund accumulation. In the "captured" value chain, owing to the change of the global competition environment, the continuous entry of suppliers and the competition effect of cheap commodities, the foundation for the existence of some previous obvious advantages that can be acquired by relying on international big purchasers has already been seriously eroded, and has been replaced with more and more obvious defects and shortcomings, which constitutes a serious threat and hazard to the industrial upgrading process of developing countries. Under GVC, as international subcontractors, enterprises of developing countries will, while molding their global brands and independently building sales terminal channels at home and abroad, meet with "the positional block" of transnational companies that control core technical patents and product standard systems and international big purchasers that control the sales terminal passages and brands of international demand market. Therefore, through the so-called "Learning by exporting", the skills of product and technical upgrading can be learned at most, and the real core technology and functional upgrading skills cannot be learned at all. For this, they held the opinion that the export-oriented development strategy based on GVC has made a huge contribution to economic takeoff during the early stage of economic development of Yangtze River Delta. However, in order to make Yangtze River Delta really become a base of advanced manufacturing and transfer from manufacturing to creation, we cannot merely rely on this development strategy. On the basis of attaching equal importance to international and domestic market, we should try to integrate the industrial relevancy and cycle system on which Chinese enterprises rely for survival and development, mold the governance structure of domestic value chain, and adjust the relationship structure among Chinese industries located in different regions, so as to lay a solid development platform for the manufacturing upgrading of the Yangtze River Delta and the integrated development of regional economy.

4 THE UPGRADING OF MANUFACTURING FROM OTHER PERSPECTIVES

Promoting the upgrading of manufacturing by developing producer services is an indirect upgrading method. Modern Producer Services is an industrial sector offering direct services to productive or commercial activities as intermediaries, including finance, insurance, accounting, R & D design, law, technical and management consultancy, transportation, telecommunication, modern logistics, advertising, marketing, brand, personnel, administration and property management. Obviously featured by high knowledge, intelligence, growth, employment and influence, it is derived from the matrix of manufacturing. Therefore, it has a natural and intrinsic industrial relevancy or interaction with manufacturing. As indicated by the research of Park and Chan (1989), the development of manufacturing can bring about that of producer services, which can, in turn, promote the upgrading of manufacturing. There is an obvious positive relevancy between them. In contrast, by arguing from an opposite perspective, Farrell and Hitchens (1990) pointed out that the lack of producer services or the inadequate price and competitiveness of producer services in a region will set back the efficiency, competitiveness and operation of local manufacturing, thus destroying the development process of this region. Some other scholars, including Zhi (2001) and Zhou(2003) analyzed this from the perspective of industrial amalgamation. As pointed out by them, with the continuous advancement of information technology revolution, the traditional boundary between services and manufacturing becomes vaguer and vaguer, and they tend to develop in an interactive and amalgamated way. Liu(2008) also held the opinion that the development of producer services can reduce the installation cost of manufacturing enterprises, thus assisting enterprises in forming their core competitiveness and forming their interaction in industrial relevancy.

5 CONCLUSION

There are a lot of research literatures of experts and scholars on the upgrading of manufacturing. All the researches into the upgrading of manufacturing, from such perspectives as industrial competitiveness, global value chain and "the indirect method of upgrading", have their respective theoretical bases and realistic significances for existence. All of them have made their respective due contribution in this regard. However, all the existing literatures neglect to study the upgrading of manufacturing from the perspective of the entire national systems, which is a macroscopic perspective that is difficult to grasp. From the emergence of institutional economics in the field of economics to the gradual perfection of new institutional economics, we should abandon those things regarded as hypothetical exogenous variables in the past, and turn these exogenous variables into endogenous ones, and study some practical problems from this perspective. We can discover some useful things from both transaction costs and institutional history, and from both the national theory of new institutional economics and its enterprise theory, so that we can study and practice the upgrading of manufacturing in a better way.

Economic development cannot do without systems, because economic development needs to be promoted by people, and people promote economic development by working out some systems. Such an objective fact as economic development is promoted by people subjectively. Therefore, it is very important to study this kind of subjective representation and some things hidden behind it. For example, in Chinese industrial upgrading, the promotion of government plays a very great role. However, in the process of decision making, the government does not always have a very thorough understanding of industrial upgrading. A very big misunderstanding has occurred in the present industrial upgrading, namely: as understood by many of us, the industrial upgrading can be realized through a transition from the low-level facet of one industrial sector to the low-level facet of another. As a result, we always provide processing sites to large transnational companies of developed countries under economic globalization. Many people understand this simple transfer from the low end of one industrial sector to the low end of another as industrial upgrading, which is, as a matter of fact, very unreasonable. Therefore, how to standardize this kind of understanding and the behaviors of the government from the perspective of systems, and how to measure the opportunity cost of this kind of upgrading from the perspective of transaction cost, will be an indispensible part of our future research.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to the financial aid of Natural Science Foundation of China (grant number:70971031, 71031003), Sanjiang University (project number: K08010),and the innovation project for the graduate students of the Business School of Nanjing Normal University(project number:10CX_003G).

REFERENCES

Chen, L. M., Wang, X., & Rao, S. Y. (2009). The comparison between the international competitiveness of Chinese and American manufacturing: Empirical analysis based on the hierarchy opinion of industrial competitiveness. *China Industrial Economy*, *6*, 57–66.

Farrell, P. N., & Hitchens, D. M. W. N. (1990). Producer services and regional development:a review of some major conceptual policy and research issues. *Environment & Planning A*, *22*, 1141–1154. doi:10.1068/a221141

Gereffi, G. (1999). International trade and industrial upgrading in the apparel commodity chain. *Journal of International Economics*, *48*(1), 37–70. doi:10.1016/S0022-1996(98)00075-0

Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy*, *12*(1), 78–104. doi:10.1080/09692290500049805

Humphrey, J., & Schmitz, H. (2002). How does insertion in global value chains affect upgrading in industrial cluster. *Regional Studies*, *36*(1), 1017–1027. doi:10.1080/0034340022000022198

Jin, B., Li, G., & Chen, Z. (2007). The status-quo analysis and enhancement countermeasures for the international competitiveness of Chinese manufacturing. *Finance & Trade Economics*, *3*, 3–10.

Li, G., Jin, B., & Dong, M. J. (2009). Basic judgment over the development status quo of Chinese manufacturing. *Review of Economic Research*, *41*, 46–49.

Lin, Y. L. (2010). The status quo of Chinese manufacturing and research into its comparison with that of foreign countries. *The Journal of North China Electric Power University*, *3*, 32–37.

Liu, L. Q., & Tan, L. W. (2006). Two-dimensional appraisal of industrial international competitiveness—The thoughts against the background of global value chain. *China Industrial Economy*, *12*, 37–44.

Liu, Z. B., & Yu, M. C. (2009). Go from GVC to NVC - The integration and industrial upgrading of the Yangtze River Delta. *Academic Learning*, *5*, 59–67.

Liu, Z. B., & Zhang, J. (2007). Forming, breakthrough and strategies of captive network in developing countries at global outsourcing system: Based on a comparative survey of GVC and NVC. *China Industrial Economy*, *5*, 39–47.

Liu, Z. B., & Zheng, J. H. (2008). *Driving the Yangtze River Delta with service industry,* (pp. 56-59). Beijing, China: The Press of Renmin University of China.

Park, S. H., & Chan, K. S. A. (1989). Cross-country input-output analysis of intersectoral relationships between manufacturing and services and their employment implications. *World Development*, *17*(2), 199–212. doi:10.1016/0305-750X(89)90245-3

Porter, M. E. (2002). *National competitive advantages*. Beijing, China: Huaxia Press.

Zhi, C. Y. (2001). The industrial amalgamation of information telecommunication industry. *China Industrial Economy*, *2*, 24–27.

Zhou, Z. H. (2003). Industrial amalgamation: The new power of industrial development and economic growth. *China Industrial Economy*, *4*, 46–52.

Zhuo, Y. (2009). The governance of global value chain---Upgrading and the performance of 5 local enterprises---The questionnaire survey and empirical analysis based on Chinese manufacturing enterprises. *Finance & Trade Economics*, *8*, 93–98.

This work was previously published in Comparing High Technology Firms in Developed and Developing Countries: Cluster Growth Initiatives, edited by Tomas Gabriel Bas and Jingyuan Zhao, pp. 134-144, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 55
UB1-HIT Dual Master's Programme:
A Double Complementary International Collaboration Approach

David Chen
IMS-University of Bordeaux 1, France

Jean-Paul Bourrières
IMS-University of Bordeaux 1, France

Bruno Vallespir
IMS-University of Bordeaux 1, France

Thècle Alix
IMS-University of Bordeaux 1, France

ABSTRACT

This chapter presents a double complementary international collaboration approach between the University of Bordeaux 1 (UB1) and Harbin Institute of Technology (HIT). Within this framework, the higher education collaboration (dual Master's degree programme) is supported by research collaboration that has existed for more than 15 years. Furthermore this collaboration is based on the complementarities of competencies of the two sides: production system engineering (UB1) and software system engineering (HIT). After a brief introduction on the background and overview, the complementarities between UB1 and HIT are assessed. Then a formal model of the curriculum of the dual UB1-HIT Master's programme is shown in detail. A unified case study on manufacturing resource planning (MRPII) learning is presented. Preliminary results of the Master's programme are discussed on the basis of an investigation carried out on the first two cohorts of students.

BACKGROUND AND OVERVIEW

Research relationships between the University of Bordeaux 1 (UB1, France) and Harbin Institute of Technology (HIT, China) exist for several years and both parties have established strong and long-term relationships with their industries over some 30 years. In the research domain on computer integrated manufacturing and production system engineering and integration, the cooperation between the University of Bordeaux 1 (IMS-LAPS: Laboratory for the Integration of Materials into Systems-Automation and Produc-

DOI: 10.4018/978-1-4666-1945-6.ch055

tion Science Department) and China started in 1993. Several Europe-China projects coordinated by UB1 have been carried out (1993-1995; 1996-1997; 1998-2002) in this domain, involving more than 7 major Chinese universities such as Tsinghua University, Xi'an Jiaotong University, Harbin Institute of Technology, Huazhong University of Sciences and Technologies, and others. More recently, the cooperation between the University of Bordeaux 1 and Harbin Institute of Technology has been strengthened to develop enterprise interoperability research activities in the Interop Network of Excellence (2004-2007) programme under the auspices of the European 6th Framework Programme for Research & Development (FP6) (European Commission, 2003b).

There is a long and strong cooperation between UB1 and HIT in research on other topics as well, including enterprise system modelling, engineering and integration. However co-operation in higher hducation was not so well-developed in the past. Consequently, it was logical to extend the existing co-operation from the research base to incorporate higher education.

Therefore, in September 2006 UB1 and HIT launched a dual master's degree programme on enterprise software and production systems. This programme relies on the know-how of HIT in computer sciences and enterprise software applications, and of UB1 in enterprise modelling, integration and interoperability research.

This joint international programme aims to train future *system architects* of production systems, with the ability to model, analyze, design and implement solutions covering organization, management, and computer science in order to improve performance of both manufacturing and service enterprises. It also aims to develop the capabilities of students to develop and grow in an international working environment particularly in China or France but also in most other countries where the themes covered by the programme are now and will continue to be vital.

The programme is organized over two years. The first year's courses are given in HIT and are concerned with industrial oriented computer sciences. The second year's courses are given in UB1 and dedicated to production management and engineering. The first two cohorts of the master's programme have successfully completed their studies and their industry internships in China and France and have obtained the Master's Degree of the University of Bordeaux 1 and the Master's Degree of Harbin Institute of Technology in September 2008 and 2009.

Table 1 gives an overview on the organization of the two year programme. All courses are presented in English, including examinations and internship defense. One characteristic is that the industry internship can be carried out in China, or in France or in any third country in the world.

Table 1. Organisation of the dual master's programme

<table>
<tr><td colspan="3">Year 1</td></tr>
<tr><td>Teaching/training</td><td>Semester</td><td>Location</td></tr>
<tr><td>Project</td><td>First</td><td>Harbin or Bordeaux</td></tr>
<tr><td>Internship</td><td>First</td><td>World</td></tr>
<tr><td>Courses</td><td>Second</td><td>Harbin</td></tr>
<tr><td colspan="3">Detail:
• Project (135h / 9 ECTS - European Credits Transfer System),
• Training in enterprise (305h / 21 ECTS),
• Algorithm and System Design and Analysis (90h / 6 ECTS),
• Database Design and Application (analysis and design) (94h / 6 ECTS),
• Software Architecture and Quality (93h/6 ECTS),
• Project Management and Software development (92h / 6 ECTS),
• Object-Oriented Technology and UML (86h / 6 ECTS).</td></tr>
<tr><td colspan="3">Year 2</td></tr>
<tr><td>Teaching/training</td><td>Semester</td><td>Location</td></tr>
<tr><td>Courses</td><td>Third</td><td>Bordeaux</td></tr>
<tr><td>Training in company</td><td>Fourth</td><td>World</td></tr>
<tr><td colspan="3">Detail:
• Modelling of industrial systems (135h / 9 ECTS),
• Production management (135h / 9 ECTS),
• Industry performance measurement (45h / 3 ECTS),
• Industry systems integration (90 h / 6 ECTS),
• Option (45h / 3 ECTS),
• Training in enterprise (450h / 30 ECTS).</td></tr>
</table>

The internship placements are mainly in companies, large as well as small/medium enterprises (SMEs), which have industrial co-operation projects with China, but not necessarily limited to that. Besides IT-oriented work, the internships are situated in the manufacturing industry sector as well as that of the services, typically as a responsible person in charge of industrial management (production, quality, and maintenance), a person in charge of design, development and implementation of software applications, a consultant, or a project leader.

COMPLEMENTARITIES UNDERPINNING THE COLLABORATION

Software Engineering and Production System Engineering

As mentioned above, this collaboration is based on the complementary strengths of UB1 (production system engineering) and HIT (software system engineering).

Considering an enterprise from the general point of view as a system providing goods and services or, from the narrower point of view of its information system, it is clear that both of these approaches relate to the fundamental philosophy of engineering. In both cases the purpose is to design an overall architecture for the system, consistent and relevant to a predefined mission. Models and simulations have a central role in both approaches.

Production system engineers view the enterprise as a system having a purpose related to a strategy. Within this purpose and strategy, performances are defined and enable the evaluation of how well the enterprise runs.

The necessity for communication and co-operation between sections within a company or between companies within a network has led to the important concept of integration. Today, the numerous forms of co-operation and the versatility they require brings into prominence the concept of interoperability that can be broadly understood as a loose integration.

Because of the complexity of the enterprise, it is always considered to relate to a *reference* (a conceptual model or reference architecture). With respect to this reference, the engineering methodologies used are supported by modelling languages and frameworks (enterprise modelling), the role of which is to enable the understanding of the structure and behavior of the enterprise. The existing diversity of languages and software supports leads to the need to analyze them in detail, in order to compare them and potentially use them together. In this perspective, a pure syntactical approach is not enough, and therefore current scientific developments in this field are related to semantics and deal with meta-models and ontology.

Furthermore, the consideration of the human-being as a component of the enterprise must always be remembered. For this reason the relation of the models with decision-making (of design and/or of management) is an important issue, whatever the approach used.

From a software engineering point of view, the need for integration can be matched through the provision and the implementation of software tools, mainly enterprise resource planning (ERP) tools. This domain then focuses on IT solutions analysis, implementation projects, IT solution performance analysis, and the identification of the usability domain and the limitation of classical methods. The ways in which the functions of the information system are integrated using such IT tools is globally understood today. Organizational challenges are also quite well known. The main outstanding issues relate to supporting the processes of the enterprise by consistently integrating the several IT solutions that have functionalities that generally cover more than is required. In this context, the capability to match the models of the enterprise (the requirements) with the models emerging from the IT solutions (the so-called space of solutions) becomes crucial. Finally, a continuing core problem is ensuring a permanent alignment of the information system and its various implemented IT solutions with the strategy of the company. Because the economic environment is dynamic, this leads, of necessity, to a policy of continuous engineering.

In summary, the two domains relate both to the design, integration and control of systems under performance conditions. In order to match the dynamic requirements and take changing constraints into account, it is necessary to continually improve the understanding of the interactions between the various models and to gather and integrate the various points of view such as organization, software, etc. In this drive to keep on improving performances, the exploitation of the complementarities between software engineering and production systems engineering is a thoroughly necessary requirement.

Enterprise Interoperability as an Emerging Topic Related To These Complementarities

Enterprise interoperability is a topic currently emerging at the confluence of software engineering and production systems engineering. It is a topic of considerable and growing scientific and technical research, fundamentally because of the considerations presented above.

Worldwide, the competitiveness of enterprises, including SMEs, will strongly depend in the future, on their ability to develop and implement massively and rapidly networked dynamic organisations. New technologies for interoperability within and between enterprises will have to emerge to radically solve the recurrent difficulties encountered - largely due to the lack of conceptual approaches - to structure and interlink enterprises' systems (information, production, decision) (European Commission, 2003b).

Today, research on interoperability of enterprise applications does not exist as such. As a result of the IST Thematic Network IDEAS (Baan, 2003), the roadmap for interoperability research emphasises the need for integrating three key thematic components, shown in Figure 1:

- software architectures and enabling technologies to provide implementation solutions
- enterprise modelling to define interoperability requirements and support solution implementation
- ontology, to identify interoperability semantics in the enterprise.

Interoperability is seen as the ability of a system or product to work with other systems or products without special effort on the part of the user/customer (Baan, 2003).

The ISO 16100 standard (2002) defines manufacturing software interoperability as the ability to share and exchange information using common syntax and semantics to meet an application-specific functional relationship through the use of a common interface. The interoperability in enterprise applications can more simply be defined as the ability of enterprise software and applications to interact usefully. The interoperability is considered to be achieved if the interaction can, at least, take place at the three levels: data, application and business enterprise through the architecture

Figure 1. Three key thematic components and their integration (Baan, 2003; European Commission, 2003b)

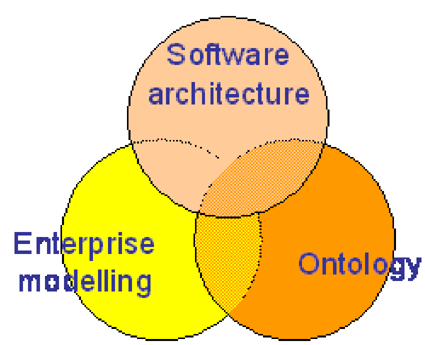
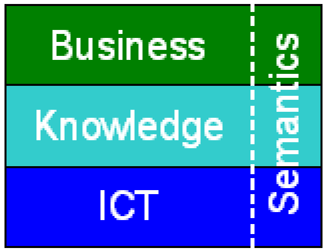


Figure 2. The three levels of interoperability (European Commission, 2003a)



of the enterprise model and taking semantics into account, as shown in Figure 2.

At the beginning of the 2000s, research in the interoperability domain in Europe was badly structured, fragmented, and sometimes overlapping unnecessarily. There was no unified consistent vision and no co-ordination between various European research centres, university laboratories and other bodies. Not only was this the case with the pure research, but it was true in the training and education areas as well. To improve this situation, two important initiatives were launched by the European Commission: Interop Network of Excellence and Athena Integrated Project (European Commission, 2003a; 2003b).

The Interop Network of Excellence and the Athena Integrated Project

Interop NoE was a Network of Excellence (47 organizations, 15 countries) supported by the European Commission for a three-year period (2003-2006) (European Commission, 2003b). This Network of Excellence aimed to extract value from the sustainable integration of these thematic components and to develop new industrially significant knowledge. Interop's role was to create the conditions of a technological breakthrough to avoid enterprise investment being simply pulled by the incremental evolution of the IT becoming commercially available.

Consequently, Interop's joint programme of activities aimed to:

- integrate the knowledge in ontology, enterprise modelling, and architectures to give sustainable sense to interoperability
- structure the European research community and influence organisations' programmes to achieve critical research mass
- animate the community and spread industrially significant research knowledge outside the network.

In more detail, the joint research activities were composed of the following work packages:

- enterprise modelling and unified enterprise modelling language (UEML): unifying for interoperability and integration
- ontologies for interoperability
- domain architecture and platforms
- domain interoperabiliy
- synchronization of models for interoperability
- model driven interoperability
- model morphisms
- semantic enrichment of enterprise modelling, architectures and platforms
- business/IT alignment
- methods, requirements and method engineering for interoperability
- interoperability challenges of trust, confidence/ security
- services/take-up towards SMEs.

Athena (Advanced Technologies for Interoperability of Heterogeneous Enterprise Networks and their Applications) was also an Integrated Project supported by the European Commission for the three-year period (2003-2006) (European Commission, 2003a).

Its objective was to be the most comprehensive and systematic European research initiative in the field of enterprise application interoperability, removing barriers to the exchange of information within and between organizations. It would perform research and apply results in numerous industrial sectors, cultivating and promoting the *networked* business culture. Research and development work was carried out hand in hand with activities conceived to give sustainability and community relevance to the work done. Research was guided by business requirements defined by a broad range of industrial sectors and integrated into piloting and training. Athena would be a source of technical innovations leading to prototypes, technical specifications, guidelines and best practices, trailblazing new knowledge in this field. It would mobilize a critical mass of interoperability stakeholders and lay the foundation for a permanent, world-class hub for interoperability.

Projects running within Athena were organized in three action lines in which the activities would take place. The research and development activities were carried out in action line A. Action line B would take care of the community building while action line C would host all management activities (European Commission, 2003a).

Concerning the R&D action line, six projects were initially defined as follows:

- enterprise modelling in the context of collaborative enterprises (A1)
- cross-organisational business processes (A2)
- knowledge support and semantic mediation solutions (A3)
- interoperability framework and services for networked enterprises (A4)
- planned and customisable service-oriented architectures (A5)
- model-driven and adaptive interoperability architectures (A6).

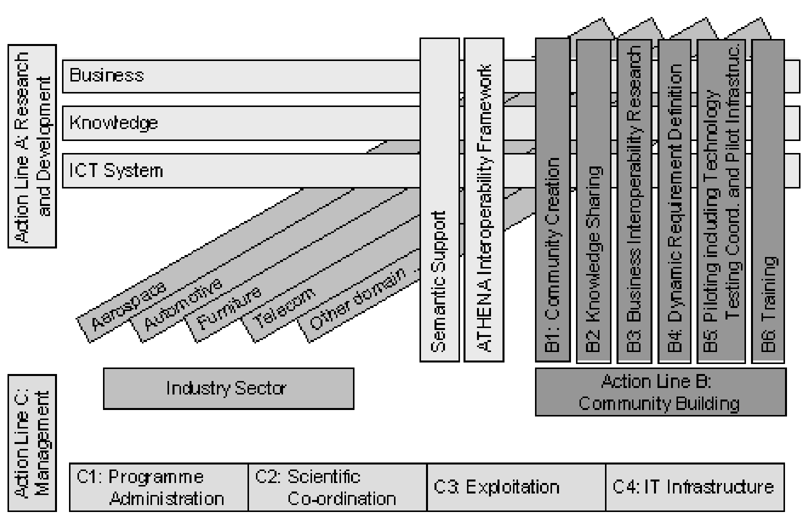
Relations between the three action lines are shown Figure 3.

Interop NoE and Athena IP have strongly influenced and contributed to research and development on enterprise interoperability in Europe and beyond. Harbin Institute of Technology was also been invited to participate in Interop NoE meetings and in the creation of the Interop Virtual Laboratory which is considered one of the important achievements of this Network of Excellence.

The Interop Virtual Laboratory (Interop-VLab)

Interop-VLab, a sustainable European scientific organization, is the continuation of the Interop

Figure 3. Interaction of Athena action lines



Network of Excellence. It aims at federating and integrating current and future research laboratories, both academic and industrial, in order to fulfil objectives that a participating organization would not be able to achieve alone. It is supported by local institutions to promote interoperability in local industry and public administration. Interop-VLab's mission includes the following:

- **Promoting the enterprise interoperability domain and acting as a reference**: establishing a sustainable organization, at European level, to facilitate and integrate high level research in the domain of enterprise interoperability and be a reference for scientific and industrial, private and public organisations
- **Contributing to the European Research Area**: contributing to solving one of the main issues of the European Research Area - the high fragmentation of scientific initiatives - by synergistically mobilizing European research capacities, enabling the achievement of critical mass by aggregating resources to match major future research challenges that would not be possible by individual organisations
- **Developing education and professional training**: promoting and supporting initiatives of European higher education institutions in the domain
- **Promoting innovation in industry and public services**: facing the industrial challenge of creating networks and synergies, Interop-VLab aims to promote and support applied research initiatives addressing innovation and the reinforcement of interoperability between enterprises, at European, national and local levels; this approach will also help to create synergy between European, national and local research programmes.

Harbin Institute of Technology is the leading partner of the China Pole of Interop-VLab. The China Pole is constituted of ten important Chinese universities spread across China. Besides research related projects, an Interop master's degree programme involving Interop-VLab members including HIT and UB1 was also planned.

FORMAL MODEL OF THE UB1-HIT DUAL MASTER'S DEGREE CURRICULUM

This section presents the details of the dual UB1-HIT master's degree curriculum. Because this programme is built on two separate disciplines and carried out in two locations in two different countries, the main challenge to its success would be the development of a deep mutual understanding of the curriculum implemented in each location and a close collaboration between the two teams, to avoid unnecessary redundancies and emphasizes synergistic complementarities. To meet this objective, a detailed and explicit representation of the curricula was necessary.

Usually university training curricula are presented in a textual form, often using tables. In general, inter-relationships between various courses and lectures tend not to be identified and/or explicitly described and considered. Sometimes this can create difficulties for students in fully understanding the relationships between component courses and their logic, and consequently in mastering the overall knowledge that they need to acquire (Alix et al., 2009).

Based on the feedback from the students after three years running on an experimental basis, it is necessary to present the master's degree programme overall curriculum in a more formal and explicit way so that both students and teachers on both sides can have a clear and unambiguous understanding of the contents of the programme and of their roles within it. Therefore, the purpose of this section is to present the formal model of the

UB1-HIT dual master's programme curriculum. Unified Modelling Language (UML) was chosen to model the lectures delivered in the two years and the possible relationships between the series of lectures in the two years. Complementarities and potential future improvements are also discussed below.

Model of Year 1 Curriculum in HIT

This section describes and model of the Year 1 curriculum carried out at Harbin Institute of Technology School of Software in China. The objective of the Year 1 training is focused on software engineering, information systems analysis and design, programming techniques and IT project management.

This curriculum is mainly organized in three modules as shown in Figure 4: Language; Science and Methodology; IT Technique.

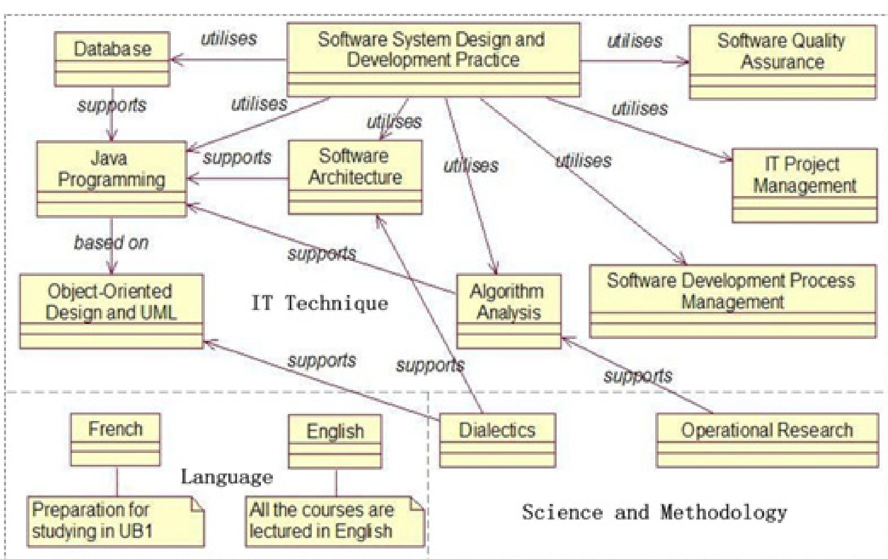
In the Language module, there are two courses, English and French.

- **English:** Because of all the courses of this joint master's programme are in English, a command of English is very important. The objective is to give the students the ability to read and write reports/papers in English, and to communicate with professors fluently, orally and aurally, in English.
- **French:** This course aims to teach the Chinese students daily French, which can help them to adapt to French daily life when they arrive in France.

The Science and Methodology module aims to teach students how to carry out scientific research, how to analyse the objects in the universe and the relationships among them. This module contains two courses, dialectics and operational research.

- **dialectics:** This course is to teach students the resolution of disagreement through rational discussion and ultimately the search for truth.
- **operational research:** This shows how to use mathematical modelling, statistics, and algorithms to develop optimal solutions to solve complex problems, improve decision-making, and make process efficiencies, to finally achieve a management goal.

Figure 4. UML model of year 1 curriculum at HIT



The IT Technique module is the main part of the first year study. This centres on software engineering. It offers a series of IT technique courses, such as databases, Java programming, etc., as well as a series of software management courses, such as software quality assurance, IT project management, etc.. In addition, there is a practical course in this module, in order to put both IT and project management knowledge into practice.

- **IT:** this set of modules aims to teach students the skills of design and implementation of IT solutions for different kinds of firm. The modules are as follows.
    - **databases:** this module focuses on how to use a relational database, including, designing a proper entity relationship model (ERM), creating correct data view based on ERM, querying data by structured query language (SQL), defining store procedure for a database, etc.
    - **algorithm analysis:** this module is an important part of broader computational complexity theory, providing theoretical estimates for the resources needed by any algorithm to solve a given computational problem: it shows how to analyze an algorithm, how to determine the amount of resources (such as time and storage) necessary to execute it, and finally achieve the goal of optimising the program.
    - **software architecture:** this module shows how to analyze, design and simulate the structure or structures of the system - the software components, the externally visible properties of those components, and the relationships between them.
    - **Java programming:** this module introduces one of the most popular programming languages: after completing this module, students should have the ability to implement an executable application and learn other programming languages by themselves.
    - **object-oriented design and UML:** unified modelling language (UML) is a standardized general-purpose modelling language in the field of software engineering: it includes a set of graphical notation techniques to create visual models of software-intensive systems; after this course, students should have the ability to use UML to design a proper software system model.
- **Management:** This set of modules contains lectures on the methodology of IT project management. The courses involve the following modules.
    - **software quality assurance (SQA):** this topic covers the software engineering processes and methods used to monitor and ensure quality: it encompasses the entire software development process - software design, coding, source code control, code reviews, change management, configuration management, and release management.
    - **IT project management:** this topic shows how to lay out the plan for an IT project, and how to realize, and anticipate and avoid the risks of failure of the IT project development: after this course, students should be able to use the methodology learned to reduce the cost of the IT project and to make the project efficient and as successful as possible.
    - **software development process management:** this module gives more details about SQA and IT project management in the development phase of a project.

- **Practical work:** This module gives students a chance to put their knowledge into practice. Students are required to manage a full IT project by themselves, from requirement analysis, system model design to software implementation, test, and then software deployment: after completing this module, students will have an overall understanding of software engineering.

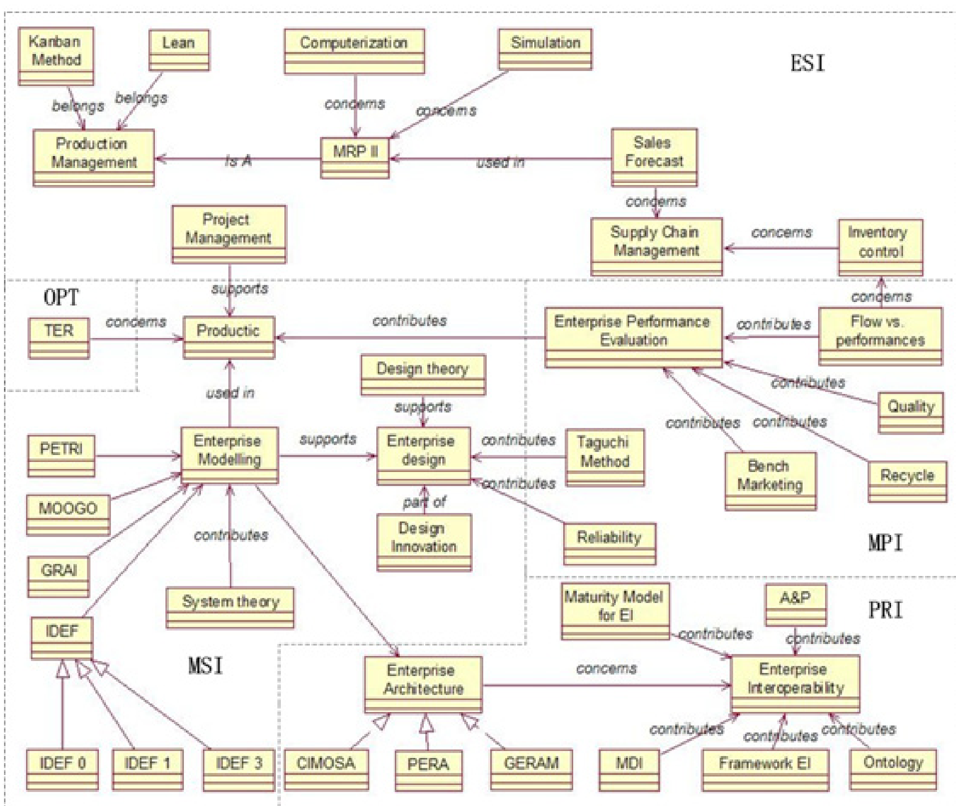
Model of Year 2 Curriculum in UB1

This section presents the model of the Year 2 curriculum at the University of Bordeaux 1 in France. The objective of this training is focused on enterprise system engineering, and in particular, enterprise modelling, production management, enterprise integration and interoperability.

The curriculum of year 2 is organised in five modules, as shown in Figure 5: MSI (industrial system modelling); ESI (industrial system management); MPI (industrial system performance); PRI (industrial system integration); OPT (option - bibliographical research work).

The MSI module is mainly concerned with enterprise modelling and design. It starts with a lecture on system theory, laying down the fundamental concepts of the systemic view of the enterprise. Then enterprise modelling focuses on GRAI (graphs of interlinked results and activities) and IDEF (integration definition) methodologies (IDEF0 function modelling, IDEF1 information modelling and IDEF3 process modelling). The MOOGO (method for object-oriented business process optimization) process modelling tool developed by the Fraunhofer Institute for Produc-

Figure 5. UML model of year 2 curriculum at University of Bordeaux



tion Systems and Design Technology (IPK) of Berlin and Petri net formal modelling are complementary to GRAI and IDEF. Productic (production science) is a lecture presenting the general problems and state-of-the-art of enterprise engineering. In parallel, design theory and innovation are presented to allow understanding of the basic concepts and principles of enterprise system design.

The ESI module focuses on production planning and control techniques with the emphasis on the MRPII method. MRPII teaching is mainly organised around an extended case study (details are given below), including (a) paper exercises, (b) game based simulation, (c) computerisation using *Prélude* software (Chen & Vallespir, 2009). Sales forecasting and inventory management methods (for example, the order point method) support both manufacturing resource planning (MRPII) implementation and supply chain management which is also another important lecture in this module. In addition, other recent methods, such as KANBAN based on JIT (just in time) and lean manufacturing, allow complementing MRPII. In parallel, project management techniques such as the PERT (programme evaluation and review technique) method are also presented.

The MPI module covers enterprise performance evaluation. Besides the Taguchi method and the reliability approach which can be related to design issues in the earlier MSI module (as shown Figure 5), a large part of the teaching is focused on quality concepts and methods. Benchmarking is also considered an important approach to improving the performance and quality of the enterprise systems and products. Another lecture is concerned with problems and solutions for recycling which is becoming more important in modern industrialised societies. Finally a game based on simulation shows how to link the flow (physical, information) in an enterprise to the performance (quality, delay), and how to act on the flow to improve the performance.

The PRI module is about enterprise integration and interoperability. Here, enterprise integration is approached principally through the use of enterprise architecture and framework modelling approaches, such as CIMOSA (computer integrated manufacturing open system architecture), PERA (Purdue enterprise reference architecture) and GERAM (generalised enterprise reference architecture and methodology). In parallel, basic concepts, framework and metrics for enterprise interoperability are also presented, because these are becoming significant new trends replacing traditional integration oriented projects. It is also noteworthy that teaching in this module is largely based on e-learning on the one hand and on the other, on seminars presented by well-known European experts in MDI (model driven interoperability), A&P (architecture & platform) for interoperability, and ontology for interoperability.

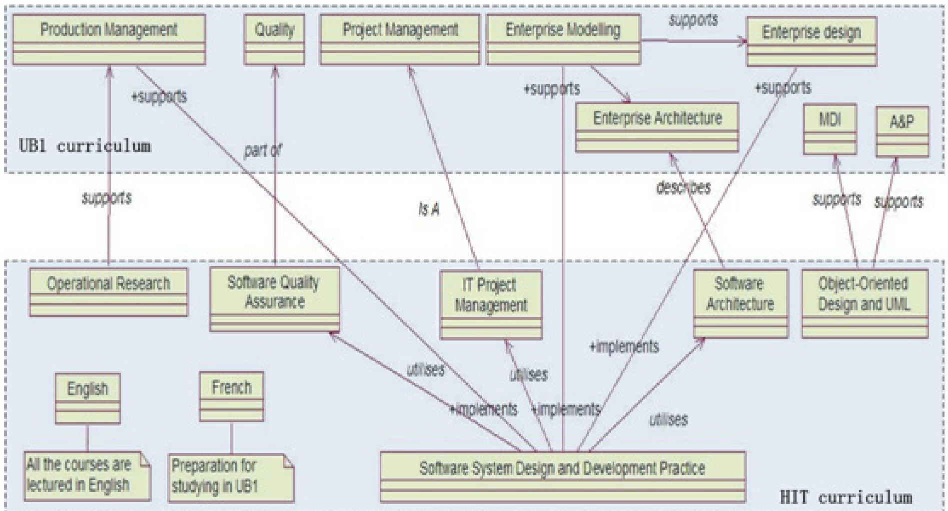
Finally the OPT module was originally designed to be a slot for optional courses. For the time being it has only one option (bibliographical research work). The students are asked to choose a subject proposed by professors and perform a bibliographical research on this. This work is done by groups of two students. Each group must write a report, present the work and answer questions in front of a jury. This work is an initiation to research work and aims at developing the capability of students to carry out bibliographical research.

Complementarities and Possible Improvements

Relationships between the courses in years 1 and 2 are tentatively identified as indicated in Figure 6. Several types of relationships are defined as follows:

- *is a* relationship: for example the IT project management lecture given in year 1 is a particular type of project management (general) studied in year 2
- *part of* relationship: the software quality assurance lecture in year 1 is part of more general quality course in year 2

Figure 6. Links between the courses of the two years



- *support* relationship: this means that one course is used as a preparation or a means for another one, such as for example software oriented design and UML that are used to develop MDI and implement A&P in year 2. Enterprise modelling techniques can also be used to model user's requirements at higher level abstraction in software system design, for example, control and information management (CIM) level in the model drive architecture (MDA) framework.

Several complementarities can be identified.

- At the global level, courses on computer science are complemented by training on enterprise and production systems. This allows HIT students to acquire supplementary knowledge to be better able to develop production system oriented software such as enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM) and others. On the other hand, UB1 students who are more familiar with industrial systems are empowered with software development skills.
- At a more detailed level and from the modelling point of view, enterprise modelling (mainly at conceptual level focusing on global system modelling) is complementary to IT oriented modelling. This is also true from the architecture perspective where enterprise architecture needs to be detailed in IT architecture and IT architecture must also be consistent with enterprise architecture.
- Both Years 1 and 2 deal with design issues. Design related lectures in year 2 (design innovation, design theory, Taguchi, reliability, etc.) provide generic design concepts and principles complementary to software design techniques learned in year 1.

At the course level, several potential improvements are envisaged as follows:

- Better coordination on the project management courses of the two years is needed. A consistent framework is necessary to position each lecture to show links and complementarities.
- More explicit relations between IT architecture and enterprise architecture must be

defined, and, in particular, the alignment between business/IT, and the consistent elaboration of IT architectures in relation to enterprise architecture.

A UNIFIED MRPII TRAINING CASE STUDY

Professional training in universities on MRPII-based production planning and control techniques as well as its implementation is one of the key issues in most of the production related master's degree programmes in France. Quite often, MRPII-based education and training do not reach a satisfactory level in university curricula. There are several reasons for this. One is the lack of production and industry concepts and experience among most master's degree level students. Another reason relates to the high conceptual character of production planning and management methods, requiring mastery of many abstract ideas, definitions and terms. The third reason is that the lectures, exercises and practical work on computers usually deal with different discrete examples, case studies and illustrations. A unified common case study allowing students to learn, understand, analyse and practise MRPII-based production planning techniques is still elusive.

In this section, an innovative and experimental MRPII training project is presented. This project was first implemented in the master's degree programme (in engineering, direction and performance of industrial systems (IPPSI)) at the University of Bordeaux 1 during academic year 2008-2009, and has been partly used on an experimental basis in the dual UB1-HIT master's programme. The characteristic of this project is to combine an MRPII game, enterprise modelling (the GRAI methodology) and software implementation within a single common case study. The objective of the project is to provide the students with a unified and consistent case study to learn MRPII-based production planning, from the fundamental concepts, through paper exercises and manual game simulation to the implementation of an MRPII-based software system. After the presentation of the principles and broad organisation of the project, we will show the various phases the students follow to learn MRPII-based production planning and control in a gradual and systematic manner. The experiences of the students obtained through formal feedback and possible improvements in the approach will also be discussed.

Description of the Case

Turbix (Centre International de la Pédagogie d'Entreprise (CIPE), 2008b) is a small company that manufactures reduction gears referenced from R1 to R8 (8 finished products). The reduction gears are composed of two types of parts, E1-E8 manufactured in the company, and P1-P5 purchased externally. The E1-E8 parts are manufactured using two types of raw materials, M1 and M2. Figure 7 shows the structure of R3.

Turbix is organised in two workshops, the machine shop to manufacture the E parts and the assembly shop to manufacture the finished products (R). Masteel and Fournix are two suppliers providing raw materials M and purchased parts P, respectively. The overall organisation and physical flow is shown in Figure 8.

Figure 7. Example: R3 product structure (Centre International de la Pédagogie d'Entreprise (CIPE), 2008a)

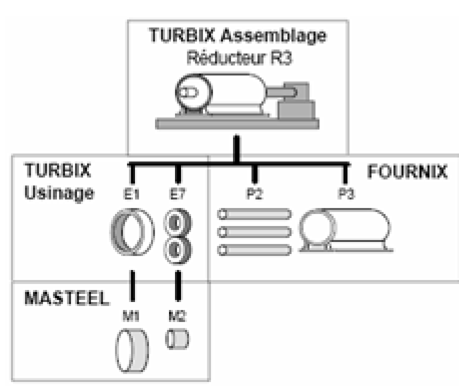
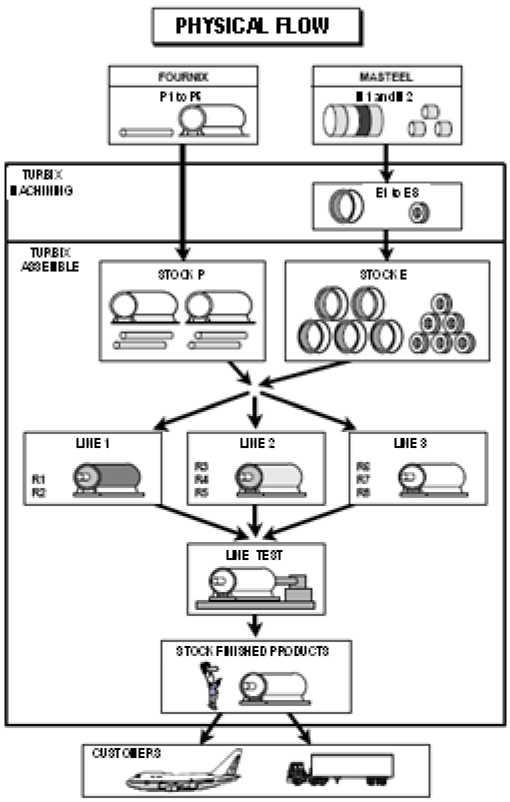


Figure 8. Organisation and physical flow of Turbix (Centre International de la Pédagogie d'Entreprise (CIPE), 2008a)



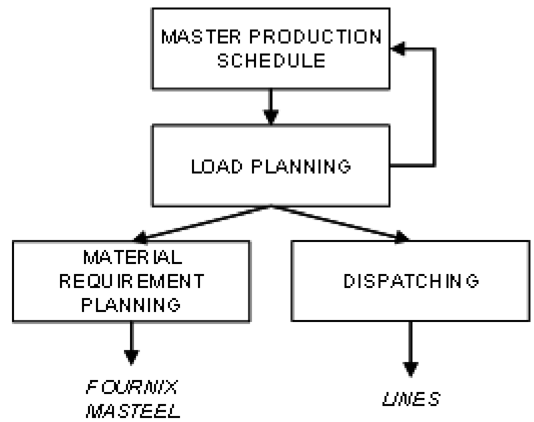
Because of different customer lead times, R1 and R2 are produced according to sales forecasts established beforehand. R3-R8 are manufactured upon firm customer orders. E1-E8 and P1-P5 are manufactured and purchased according to the needs for R1-R8 production. M1 and M2 are purchased according to the needs for E1-E8 production.

On the basis of this physical organisation, the architecture of the production management implemented in Turbix is presented Figure 9.

First Component: The Manufacturing Resource Planning (MRPII) Game

The objective of the MRPII game (Centre International de la Pédagogie d'Entreprise (CIPE), 2008b) is to allow a group of participants to discover for

Figure 9. Turbix management architecture



themselves how the MRPII method works and what are the steps one must follow to implement MRPII software in a company. Participants using this game can plan the production and purchasing orders using the MRPII technique, and simulate the execution of planned orders through various functions of the company - commercial service, manufacturing service, inventory/stocks, purchasing service. etc. During the simulation, each participant takes a precisely defined role/responsibility.

In detail, the game allows students

- to understand the structure and functioning of the existing production system
- to plan the master production schedule (MPS) for the finished products and draw up the material requirement planning (MRP) for parts E and P
- to calculate load and perform load levelling
- and finally to simulate the functioning of the production system over a period of two months, all consistent with the management architecture in Figure 9.

Second Component: The GRAI Methodology

The GRAI methodology (Vallespir & Doumeingts, 2006) was developed at the Department for

Automation and Production Science/Graphs of Interlinked Results and Activities (LAPS/GRAI) of the Laboratory for the Integration of Materials in Systems (IMS) at the University of Bordeaux 1. This methodology sets out to model, analyse and design the decision-making sub-systems of a production management system. The method consists of

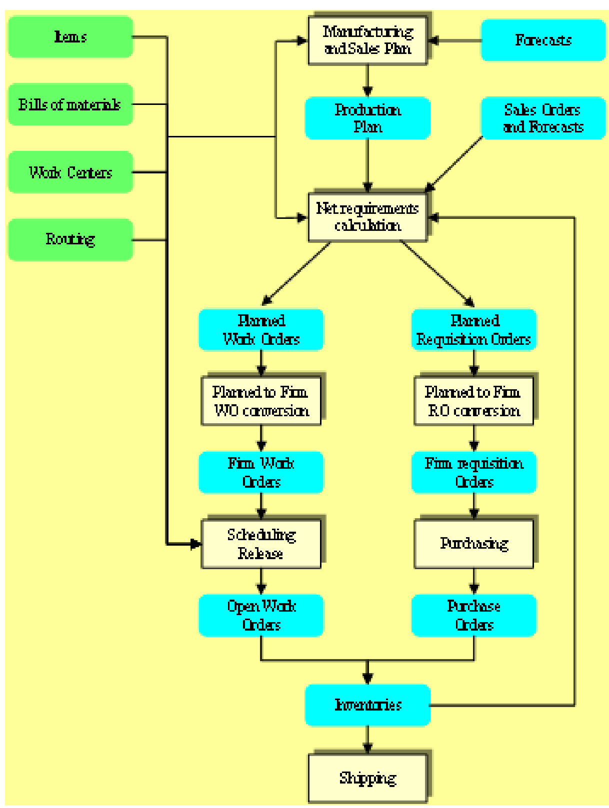
- a conceptual reference model defining the set of fundamental concepts
- modelling formalisms, and
- a structured approach.

The GRAI methodology is used in the project to model and analyse the existing production system of Turbix, to detect its potential inconsistencies and to design a new improved system.

Third Component: The Prélude Production MRPII Software

Prélude Production is an MRPII compliant software developed for professional training and teaching purpose (Centre International de la Pédagogie d'Entreprise (CIPE), 2008a). Its user-friendly interface allows students to learn how to manipulate MRPII software in a gradual way. This software is used in the project to computerise the production planning and management activities in Turbix Company. After the implementation of Prélude Production in the company, it is used to plan and control the daily production activities. It is also used together with the game to perform a simulation. Figure 10 shows the main functions of the Prélude Production software.

Figure 10. Main functions of Prélude Production software (Centre International de la Pédagogie d'Entreprise (CIPE), 2008a)



The Programme and the Implementation of the Project

In this section, we present the programme for the project and its organisation and implementation. The project is carried out by the students over several months. Two groups of students are formed, each group of about 10 students. Figure 11 gives the overall logic of the project.

Initialisation Phase

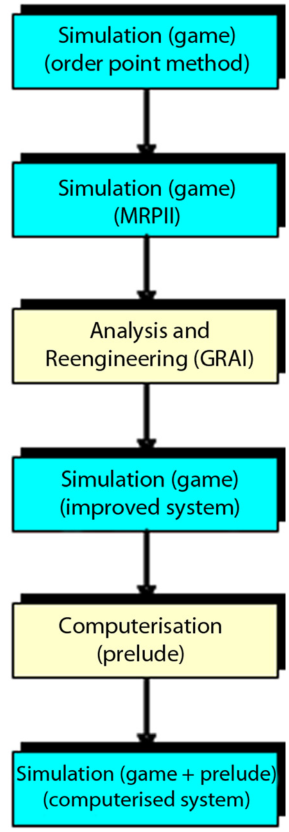
To start the project, the objective, the organisation and time table, as well as the expected results at the end of each phase are presented to the two groups of students.

Playing the Game

The next phase aims to show the students how to carry out the planning and simulation without the MRPII software tool. The objective is to allow students to develop a better understanding the basic concepts and techniques of the MRPII calculation, and, at the same time, a thorough understanding of the existing Turbix system.

The game is played over one day and a half. At the beginning the students use the traditional inventory management technique (order point method) to manage the Turbix system for a one month (January) period. Then they are asked to migrate to the MRPII technique. Manual MRPII calculation is done to plan all the orders needed for the finished products (R1-R8), and parts (E1-E8) and (P1-P5). Load calculations on the four lines in the assembly workshop (the L1-L3 assembly lines, and the L4 test line) are carried out in order to validate the master production schedule (MPS). The MRPII simulation is launched on a day-by-day basis for the duration of the next month (February), for managing the production activities (purchasing, manufacturing and assembly) and the management activities (order release, production follow-up, inventory, orders close-up, etc.).

Figure 11. Overall logic of the project



Existing System Analysis

After the MRPII game, the students are asked to analyze the functioning of the existing production system based on their knowledge and experience gained during the game. The GRAI method, using GRAI grid and nets, is used to model the decision-making structure of the existing project management (PM) system. Based on the model of the existing system, GRAI rules can be applied to detect possible inconsistencies. If inconsistencies are found (for example, a bad decision horizon or faulty period values), the students will propose necessary corrections to the existing system in order to improve its functioning.

Simulation of Improved System

After the analysis and possible redesign of the production system in Turbix, the students then play the game again. The game simulation is done on the new system, having implemented the set of suggested corrections and modifications to the existing system. For example, one of the possible suggestions that might be proposed by the students is to adjust the value of the planning horizon in the MPS and the MRP levels to allow an improved co-ordination between them.

Implementation of Prélude Production Software

During this phase, the students are asked to computerise the production planning and control activities in Turbix using Prélude Production software. For this task, the students are divided into small groups (2 students per group per computer). Firstly, the students need to make a compilation of all the relevant technical data (bill of materials, routings, items and workstations) and put them in an appropriate form to be entered in the computer. Then a small scenario (using a number of sales forecasts and firm orders) is given to students to allow them to test the Prélude Production implemented for Turbix. This tends to be a very interesting task because the students need to find errors they may have made during the data collection.

MRPII Software Based Simulation

After the validation of these test results, the simulation can begin. During the simulation, the students are asked to perform the same activities they did during the game but this time using the MRPII software.

This phase allows students to compare the two simulations, the game simulation without computer aid and the simulation with the MRPII software (Prélude Production).

DISCUSSION AND REMARKS

The experimentation carried out among students of the master's class in 2008 showed strongly the student interest and feasibility of the project. The main added values of the project were found to be the following:

- The project allowed the students not only to learn MRPII concepts and techniques, but also to practise the MRPII-based production control in a concrete and unique case study. The students could evaluate and compare the problems, difficulties and benefits at the different stages of the project using the same case.
- The game played before the computerisation stage allowed the students to take an active part in the activities of the enterprise as if they were actors in the company, thus putting them in a situation similar to that in the real enterprise
- The use of the GRAI method before computerisation allowed the detection of possible inconsistencies in the system. The benefits are to show the usefulness of enterprise modelling to improve company performance, and to computerise a re-engineered system after the correction of inconsistencies.
- The project showed that computerisation of production management is not only a matter of software. Before introducing an MRPII package in a company, it is necessary to analyse and re-engineer the existing system to make it consistent, to have the appropriate technical data, to define the most suitable parameters for the software, etc..

This project has contributed to improving MRPII-based production management training courses in French universities by providing a unified case study framework which covers the various types of exercises (understanding fundamental concepts, paper-based MRPII planning and manual simulation, enterprise modelling/analysis and re-engineering, computerisation, MRPII software-based simulation). One of the improvements planned for the near future is the reinforcement of the use of the GRAI methodology in the second phase of the project. It will also be necessary to investigate ways of increasing the time horizon of the simulation (from two months, possibly to 6 months or preferably one year). The extension of the time horizon will allow simulation of a long term production plan and the incorporation of some strategic production management decisions.

PRELIMINARY ACHIEVEMENTS AND ASSESSMENT

This section describes the results of the two first cohorts of students on the master's programme, and the feedback received from them. The students are asked to give a personal overview and overall appreciation on the content of the programme as well as the difficulties encountered in studying and comprehending each year and the benefits expected at the end of the programme. Finally the students who have earned all the European Credit Transfer System (ECTS) credits at the end of the second year of the programme and are eligible for the dual master's degrees, are asked for a professional perspective/discussion on the advantages of the programme and degree.

The first cohort of students, class 2008, had thirteen students, twelve Chinese and one French. The second cohort, class 2009, had fourteen students in year 2, ten Chinese and four French. The third cohort, class 2010, has fifteen students in its year 1, ten Chinese and five French. The relatively low number of French students, although growing, is probably because the predisposition to go abroad for study is weak in France and the students who go aboard tend to be pioneers.

The employment opportunities for the graduates are in both manufacturing and service companies. Graduates can become managers and more specifically production, quality, or maintenance managers, R&D engineers and managers, consultants, project coordinators and managers in the general domain of implementing enterprise software applications (such as ERP, SCM, PLM and many others) in large companies and in SMEs. If, as would seem likely, the internship is a springboard to employment, another employment opportunity is in research teams and projects in academic institutions. Indeed, in 2008 eight students did their final internship in an academic or research laboratory, three in France and five in other European countries. In 2009, ten students have chosen research internships, five in French laboratories and five in other European ones.

Survey of the Opinions of the Students

In December 2008, a questionnaire was sent to all the students of class 2008 and class 2009. The objective was to obtain an evaluation of the programme, taking into account the student's difficulties, the facilities and their expectations before, during and after the programme, and to obtain feedback on the professional experience gained after the two periods of internship by the two cohorts. A simplified view of the questionnaire used is presented below.

Questionnaire Used

1. Position, name and address of the company or university?
2. Position of the internship activity (daily job) in the company?
3. Competencies before year 1, before year 2 and at the end of year 2
4. Difficulties met and facilities provided during the first and the second year of the programme?
5. Advantages and disadvantages offered/encountered in relation to the double com-

petency, EM (enterprise modelling) and IT (information technology), of the programme ?
6. Thoughts about the continuity between Harbin and Bordeaux
7. In your daily job do you use the double competency (if not, which one do you use), advantages offered by IT / EM knowledge in your job?
8. Differences and similarities between the form and operation of the internship in Harbin and in Bordeaux?
9. Is the double competency an advantage in finding a job or PhD position?

Seven students of class 2008 replied, two of them employed in private companies, three PhD students, and two looking for a job or further training opportunities. Twelve students of class 2009 replied.

Results from Class 2008

- **Competencies**: Most students (5) had low or only a fundamental level of software programming skills, and some students (2) had no software domain knowledge but principally mathematics or control theory and engineering respectively before the programme year 1. At the end of the first year, almost all the students (6) had achieved competency in software engineering, especially software architecture, software development, Java and databases. At the end of two years, most students believed they had acquired (i) knowledge about enterprise modelling and production management, (ii) knowledge about enterprise modelling methods like GRAI and IDEF, (iii) deep understanding of SCM, quality assurance and performance measurement, and (iv) knowledge through the bibliographic research work in academic fields like ontology and interoperability. After the full programme, most of the students agreed that they had made progress in the English and French languages.
- **Difficulties and facilities**: During the first year of the programme, most problems came from language misunderstanding which made some courses difficult to assimilate. Three students thought that the courses were heavy even though they had a good studying environment (2 of them lacked knowledge and experience in software engineering). In the second year of the programme, 4 out of the 7 students who responded thought that topics such as interoperability and service oriented architecture were too conceptual and difficult to comprehend. With insufficient background knowledge of practical enterprise cases, models that are abstract and connections between these models are hard to understand.
- **Double competency statement**: Students have acquired knowledge of software development and enterprise modelling by the end of the two years. They have good knowledge of how IT works in the enterprise and also a good understanding of business processes which can help them to find the right technology when they design an enterprise management system. In their daily jobs five students out of the seven use this double competency. IT knowledge is used directly and regularly by persons in employment in companies while the PhD students use IT to implement programs to prove, analyze and show their research results. For those in employment, enterprise management knowledge supports their understanding of the framework and architecture of the issues they work on and supports the design of solutions in their daily work.
- **Teaching specificities**: The teaching in the IT domain tends to be considered more theoretical while the teaching in the EM domain is considered more practical because

of the game-based simulations and exercises that can be seen as playing realistic roles. The enterprise games are also considered a useful tool to explore a particular context and have special values because most these games tend to be team-oriented. In Harbin the internship takes place at the same time as the course. Consequently, students have a complete project in which they use IT technology to carry it out. An advantage, according to the students, is that they can go deeper into detail through asking for information from the teachers but that sometime this becomes too closely detailed to form a proper overall view. In Bordeaux, the internship has a specific period and the subject in question is sometimes disconnected from the course even if that subject deals with management. This requires more individual initiative and creativity because the students can feel alone in confronting their problems even if they can ask their teacher. But it is considered a strong advantage that the students are totally immersed in the company.

Results from Class 2009

- **Competencies**: At the beginning of year 1, 9 of these students had competencies in software engineering: operating system, data structure, databases, IT project management, software quality assurance and some popular development languages such as Java, C++ and.NET. One student had specialized in automatic control and another had knowledge linked to mechanical engineering and production management. Thus the competencies were much more diverse than in the previous cohort. Before year 2, most of the students (9) had improved their programming skills as software engineers. By then they had more experience in programming and project management, and knowledge of advanced databases, algorithm, software architecture and so on. They had also improved communication skills, with a good level of French and fluent English. The other two students had acquired knowledge of programming using Java, database design, and IT project management. At the end of semester 4, all the students had gained knowledge in enterprise computing and engineering, including production management, enterprise modelling, and quality management.
- **Difficulties, facilities**: Like the previous cohort, the first difficulty cited is language. The second arises from the fact that students are not au fait with the production environment and so concepts relative to an enterprise are difficult to comprehend, the concept of interoperability for example is understood, but the finer details are not, and while the model-driven architecture and enterprise modelling methods are readily learned, the lack of experience makes their use far from obvious. All the students complained about the schedule of the course, with too many courses planned in too short a period and too many different types of knowledge to be learned in different areas/domains.
- **Double competency statement**: Despite difficulties, students agreed that they had acquired a double competency. Not only did they know how to do programming but also understood how an enterprise works using IT technologies. The background of one domain was felt to be a great help when working in the other. The dual competency provides more choices for a future career. Even if it is not easy to re-orient one's mind from the software view to the enterprise view, they were confident that they would be able to bring these views together in the future.

- **Teaching specificities**: In Bordeaux, there are more games-based training exercises as opposed to the programming practicals in Harbin. As regards the internship, they did not find major differences between Harbin and Bordeaux. In the first year, the goal was to develop software systems, and students worked directly from the analysis, and then designed the system and wrote the code. In the second year, students needed to read the materials about the production system to gain a holistic understanding of the subject.

Remarks

The dual master's degee programme represents a good challenge for all the responding students because of the challenging multidisciplinary and cross-domain training during the two years. The students also became very aware of the interests and needs of companies which are very close to the topics and subjects dealt with in the programme.

Needs Expressed During the Internship (So-Called, M2)

This section provides the analysis of the internship of year 2 of class 2008, because internships of class 2009 are only in progress. As mentioned earlier, in 2008 five students did their internship in private companies. One internship topic concerned pure management issues, while all the others combined IT problems and the use of enterprise modelling methods to analyze and model enterprise systems.

Topic Relative to Management Only

A well known large company in the domain of material construction had proposed a study on pricing strategy because the market is becoming more and more competitive, and the pricing strategy must be adjusted to take into account product turnover, life cycle phase and other dynamic variables. This study focused on the analysis and comparison of the commonly used pricing strategies: premium pricing, value pricing, cost/plus pricing, competitive pricing and penetration pricing. The internship project led to the proof that the value strategy was the best strategy for new products and high-end products, but that for all other products, the competitive strategy was shown to be the best one. This conclusion has enabled the company to improve customer loyalty, keep market share and make expected profit (Jia, 2008).

Internship Studies Involving Combined Topics

One study led to the analysis of the possibilities of applying data mining techniques in cross-selling to increase the overall sales of a company specialising in material construction. The study elaborated a process methodology based on data mining software, and described the way to build mining models to do cross-selling analysis. The student described how to write associative prediction queries, integrate these queries into a Web cross-selling application and then discussed the architecture of a web application with data mining predictions (Li, 2008).

Another subject proposed by the same company concerned the exchange of data between servers of 120 commercial agencies which constitute the company. The objectives of the company were to (i) find a solution which could monitor the servers, (ii) analyse their performance, and (iii) predict potential problems and inform the system administrator in advance. Furthermore the company needed software to help the administrator do his daily work, in verifying the backup machine and the working situation of the servers, and other tasks. The mission of the student was to choose a correct solution to satisfy the company's needs and then design and implement the architecture on the existing system (Yang, 2008).

A third combined topics study related to a small company specializing in internet search engines. The main challenge of the company was to offer to internet users the relevant information about an enterprise, a product or a service. For this task the search engine limits the referencing to the web site of the enterprise in order to have consistent and precise information. Blog, personal page and forum pages are avoided. The student participated fully in the whole project, from the requirement analysis phase to the development phase, including learning and using specific languages, technologies, etc. The student acted as an actor in the project but also as project manager during the development phase (Wang, 2008).

The fourth internship was carried out in a large French worldwide company, in the Oracle project pole. The student worked on the ERP technology taking into account the requirements of a specific customer that is a public sector administration. The company maintains the Oracle IT system for the administration. The student worked on the purchase order process from the demand, to the invoice payments, including the orders and receipts. This allowed him to study the complete acquisition workflow, and introduce some new concepts of finance and accounting, and new concepts and ontology in the financial area (Fausser, 2008).

Analysis of the Topics by the Private Companies Involved

Information on projects carried out by the students during their internships in private companies was collected through the report they gave at the end of year 2 of their programme. Three subjects out of the five were proposed by one enterprise. This indicates the difficulty of finding industrial internship in France, mainly due to the language barrier. French companies find the double competency of the student very interesting but most are not prepared to integrate students who do not speak French into their company. A second important conclusion is that most of the topics (4 out of 5) required a double competency, and in those cases the students successfully applied IT techniques to improve the performance of those companies.

CONCLUSION

This chapter presents an international collaboration between the University of Bordeaux 1 and Harbin Institute of Technology. This collaboration is characterised by the fact that it is based on:

- a long-term strategy of both institutions (UB1 and HIT) to develop sustainable co-operation in the domain of interoperability which is considered a priority subject on both sides
- the two competencies, in UB1, enterprise modelling and interoperability, and production system sciences, and in HIT, computer sciences and software engineering, are complementary in the development of R&D and in this education programme
- the combination of research activities and education/training allow benefits to flow from the latest advances in research in enterprise software application interoperability (such as the European Union R&D projects, Athena, Interop, and others).

This collaboration model has considerable potential to be duplicated and extended to other universities and other countries.

The formal UML model to represent the joint master's programme curriculum allows explicit identification of all elementary lectures and the possible relationships between the lectures and modules. We believe that this formal modelling approach can help students to better understand the training curriculum and lead to an improved quality of education. Furthermore it also allows the teachers involved to check the overall consistency of the curriculum, to better coordinate and organise their lectures, to avoid unnecessary redundancies and overlapping coverage, introduce possible contractions and bring out synergies and complementarities.

The feedback on the students' experience of the dual master's degree programme shows that it responds to real business needs and concerns. Even with the language barrier, more companies are becoming interested in students with the double competency. For the students, even though the programme is difficult to assimilate during the two years because of its breadth and density, they are satisfied at the end because they have come to understand the crucial impact of IT on enterprise performance.

ACKNOWLEDGMENT

The authors thank Zhiying Tu and Zhenzhen Jia (University of Bordeaux 1) for their contribution to this chapter.

REFERENCES

Alix, T., Jia, Z., & Chen, D. (2009). Return on experience of a joint master programme on enterprise software and production systems. In B. Wu & J.-P. Bourrières (Eds.), *Educate adaptive talents for IT applications in enterprises and interoperability. Proceedings of 5th China-Europe International Symposium on Software Industry Oriented Education* (pp. 27-36). Talence, France: University of Bordeaux.

Baan, A. Z. (2003, March). *IDEAS roadmap for e-business interoperability: Interoperability development for enterprise application and software – roadmaps (IST-2001-37368)*. Paper presented at the e-Government Interoperability Workshop, Brussels, Belgium.

Centre International de la Pédagogie d'Enterprise (CIPE). (2008b). *Jeu de la GPAO* (gestation de la production assistée par ordinateur). Retrieved October 30, 2010, from http://www.cipe.fr

Centre International de la Pédagogie d'Entreprise (CIPE). (2008a). *Manufacturing resourses planning software package*. Retrieved October 30, 2010, from http://www.cipe.fr

Chen, D., & Vallespir, B. (2009). MRPII learning project based on a unified common case-study: Simulation, reengineering and computerization. In B. Wu & J.-P. Bourrières (Eds.), *Educate adaptive talents for IT applications in enterprises and interoperability. Proceedings of 5th China-Europe International Symposium on Software Industry Oriented Education* (pp. 233-240). Bordeaux, France: University of Bordeaux.

Chen, D., Vallespir, B., & Bourrières, J.-P. (2007). Research and education in software engineering and production systems: A double complementary perspectives. In B. Wu, B. MacNamee, X. Xu, & W. Guo (Eds.), *Proceedings of the 3rd China-Europe International Symposium on Software Industry-Oriented Education* (pp. 145-150). Dublin, Ireland: Blackhall.

Chen, D., Vallespir, B., Tu, Z., & Bourrières, J. P. (2010, May). *Towards a formal model of UB1-HIT joint master curriculum*. Paper presented at the 6th China-Europe International Symposium on Software Industry Oriented Education, Xi'an, China.

European Commission. (2003a). *ATHENA - Advanced Technologies for Interoperability of Heterogeneous Enterprise Networks and their Applications: Integrated project proposal. European 6th Framework Programme for Research & Development (FP6-2002-IST-1)*. Brussels, Belgium: European Commission.

European Commission. (2003b). *INTEROP - interoperability research for networked enterprises, applications and software, network of excellence, proposal part B. European 6th Framework Programme for Research & Development*. Brussels, Belgium: European Commission.

Fausser, J. (2008). *Maintenance of Oracle e-business suite V11 (Internal report of Internship of M2)*. Talence, France: University of Bordeaux.

International Organization for Standardization (ISO DIS 16100). (2000). *Manufacturing software capability profiling - part 1: Framework for interoperability* (ISO TC/184/SC5, ICS 25.040.01).

Jia, Z. (2008). *Pricing strategy for Point P based on analysis and comparison of commonly used pricing strategy (Internal report of Internship of M2)*. Talence, France: University of Bordeaux.

Li, Z. (2008). *Data mining applied in cross-selling (Internal report of Internship of M2)*. Talence, France: University of Bordeaux.

Vallespir, B., & Doumeingts, G. (2006). *The GRAI (graphs of interlinked results and activities) method*. Talence, University of Bordeaux 1: Interop Network of Excellence Project tutorial.

Wang, Y. (2008). *Improvement and extension of professional search engine (Internal report of Internship of M2)*. Talence, France: University of Bordeaux.

Yang, W. (2008). *Monitoring work situation of servers in Saint-Gobain Point P (Internal report of Internship of M2)*. Talence, France: University of Bordeaux.

KEY TERMS AND DEFINITIONS

Dual Master's Degree: A master's degree programme involving at least two different universities / institutions from two different countries, and allowing students to obtain two degrees from the two institutions.

Enterprise Modelling: Representing the enterprise in terms of its structure, organisation and operations according various points of views (technical, economic, social and human).

Interoperability: A property referring to the ability of diverse systems and organizations to work together (inter-operate). The term is often used in a technical systems engineering sense, or alternatively in a broad sense, taking into account social, political and organizational factors that impact system performance.

Production Management: A set of techniques for planning, implementing and controlling industrial production processes to ensure smooth and efficient operation. Production management techniques are used in both manufacturing and service industries.

Software Engineering: A profession dedicated to designing, implementing, and modifying software so that it is of higher quality, more affordable, maintainable and rapid to build.

This work was previously published in Software Industry-Oriented Education Practices and Curriculum Development: Experiences and Lessons, edited by Matthew Hussey, Bing Wu and Xiaofei Xu, pp. 57-81, copyright 2011 by Engineering Science Reference (an imprint of IGI Global).

Section 5
Organizational and Social Implications

This section includes a wide range of research pertaining to the social and behavioral impact of Industrial Engineering around the world. Chapters introducing this section critically analyze and discuss trends in Industrial Engineering, such as participation, attitudes, and organizational change. Additional chapters included in this section look at process innovation and group decision making. Also investigating a concern within the field of Industrial Engineering is research which discusses the effect of customer power on Industrial Engineering. With 13 chapters, the discussions presented in this section offer research into the integration of global Industrial Engineering as well as implementation of ethical and workflow considerations for all organizations.

Chapter 56
Process Innovation with Ambient Intelligence (AmI) Technologies in Manufacturing SMEs:
Absorptive Capacity Limitations

Kathryn J. Hayes
University of Western Sydney, Australia

Ross Chapman
Deakin University Melbourne, Australia

ABSTRACT

This chapter considers the potential for absorptive capacity limitations to prevent SME manufacturers benefiting from the implementation of Ambient Intelligence (AmI) technologies. The chapter also examines the role of intermediary organisations in alleviating these absorptive capacity constraints. In order to understand the context of the research, a review of the role of SMEs in the Australian manufacturing industry, plus the impacts of government innovation policy and absorptive capacity constraints in SMEs in Australia is provided. Advances in the development of ICT industry standards, and the proliferation of software and support for the Windows/Intel platform have brought technology to SMEs without the need for bespoke development. The results from the joint European and Australian AmI-4-SME projects suggest that SMEs can successfully use "external research sub-units" in the form of industry networks, research organisations and technology providers to offset internal absorptive capacity limitations.

DOI: 10.4018/978-1-4666-1945-6.ch056

INTRODUCTION

Through case study research, this chapter discusses some of the challenges Small and Medium Enterprises (SMEs) in the manufacturing sector face in identifying and adopting Ambient Intelligence (AmI) technologies to improve their operations. Ambient Intelligence technologies are also known as Pervasive computing or Ubiquitous computing, and we include the descriptions of these terms when we refer to AmI technologies. Our study includes case studies of three Australian SMEs and a comparison with similar application requirements in a German SME manufacturer. The outcomes of the study are likely to be applicable to small firms in many nations.

The 1980s and 90s saw the operations of many large manufacturers revolutionized by the introduction of process and technological innovations (Gunasekaran & Yusuf, 2002). While there have been uneven adoption rates in smaller businesses and across different nations (Chong & Pervan, 2007; Oyelaran-Oyeyinka & Lal, 2006) it is clear that technological innovations such as Electronic Data Interchange, Business Process Re-engineering, Enterprise Resource Planning and robotic automation, amongst others, have played key roles in increasing manufacturing productivity. At the beginning of the twenty first century this transformation continues. Ambient Intelligence (AmI) technologies are being positioned as the next performance and productivity enhancing purchase for manufacturers, and a potential means for manufacturers in developed nations to counter perceived threats from lower labour cost countries (Kuehnle, 2007).

Thus, the key objectives of this chapter are to consider potential applications of AmI technologies in Australian SME manufacturers, and discuss the opportunities and shared challenges faced by such firms in adopting these technologies. In doing this we will compare different levels of absorptive capacity and technological readiness in Australian firms, seeking possible reasons for similarities and differences in their comparative technology adoption processes. The chapter also examines the role of intermediary organisations in alleviating these absorptive capacity constraints.

Our overarching research question is: "Can external intermediaries overcome absorptive capacity limitations in SMEs seeking process innovation through the application of AmI technologies?" In order to understand the issues surrounding this problem, a brief overview of ICT (Information and Communication Technologies) adoption in manufacturing and an explanation of Ambient Intelligence (AmI) technologies are provided in the following section. Following that we examine the role of SMEs in the Australian manufacturing industry plus the impacts of government innovation policy and absorptive capacity constraints in SMEs in Australia.

BACKGROUND

ICT Adoption for Business Performance Improvement

Brown and Bessant (2003) described the global manufacturing environment developing in this new century as an increasingly competitive landscape, characterised by on-going demands for improved flexibility, delivery speed and innovation. A frequently occurring element in manufacturers' responses to these pressures is the implementation of increasingly sophisticated ICTs. The benefits of incorporating ICTs on business responsiveness have been identified as: more effective and more efficient information flows; assisting in value-adding improvements for current processes; greater access to efficiency enhancing innovations throughout the value chain (Australian Productivity Commission, 2004); and the ability to access world markets through e-commerce (Kinder, 2002).

ICT adoption has been considered worth the risk, given the competitive pressures placed on business to keep pace with technology. For example, in Australia, the uptake of ICTs has increased dramatically towards the later part of the 90's and into the 21st Century. Reports show that in 1993-94, 50 per cent of firms used computers with 30 per cent having internet access; by 2000-01 these figures had increased to 85 per cent and 70 per cent respectively (Australian Productivity Commission, 2004). Recent figures (Australian Bureau of Statistics, 2009) reveal that almost all Australian SMEs use ICTs, and 96% of them access the internet through a broadband connection. One of the latest developments in the application of ICTs to business improvement is that of Ambient Intelligence (AmI) technologies. The objective of AmI is to broaden and improve the interaction between human beings and digital technology through the use of ubiquitous computing devices. By using a wider range of simplified interaction devices, ensuring more effective communication between devices (particularly via wireless networks) and embedding technology into the work environment, AmI provides increased efficiency, more intuitive interaction with technology (Campos, Pina, & Neves-Silva, 2006) and improved value and productivity (Maurtua, Perez, Susperregi, Tubio, & Ibarguren, 2006).

Ambient Intelligence (AmI) Technologies

Existing literature (Kopacsi, Kovacs, Anufriev, & Michelini, 2007; Li, Feng, Zhou, & Shi, 2009; Maurtua et al., 2006; Vasilakos, 2008; Weber, 2003) points to the co-existence of three features in any AmI technology: ubiquitous computing power, ubiquitous communication and adaptive, human-centric interfaces. Regardless of arguments about terminology and definitions (the terms "pervasive computing" and "ubiquitous computing" are in common use in the US, while "ambient intelligence" is favoured in the EU), these technologies are already commonplace. The beep signalling the automatic deduction of a road toll from your account as your car passes under a toll gate is one aspect of an AmI technology known as Radio-Frequency Identification (RFID). RFID technology is having an impact in many industries, some of which are not normally associated with high levels of ICT adoption. For example, during 2006, in NSW alone (one of the 7 states and territories within mainland Australia), more than 1.2 million head of cattle were automatically tracked from farm to saleyard to abattoir as their RFID ear tags passed through RFID sensor gates (NSW Farmers Association, 2007).

In addition to increasing process speed and efficiency, AmI technologies have the potential to provide tracking of employee and customer activity. While concerns about the impact of technology upon power relations in the workplace are not new (Zuboff, 1988), the characteristics of AmI technologies present new challenges to worker privacy, informed consent and dignity. AmI technologies may, intentionally or not, dramatically increase employee surveillance and monitor consumer activity over the entire product life cycle. This potential and the very nature of 'ubiquitous computing' raises important ethical issues. Proposals to use RFID tags to track sufferers of Alzheimer's disease (Caprio, 2005) and children provide examples of the ethical dilemmas AmI technologies can present. While these issues are beyond the scope of this chapter, we suggest Cochran et al (2007) for a review of ethical challenges associated with RFID.

Social factors associated with the introduction and implementation of AmI technologies may be exacerbated in small and medium businesses. In addition to concerns shared with corporate workers, such as disquiet about their personal data potentially being sold to marketing groups and anxiety about the security of the information gathered, members of small businesses are particularly prone to AmI's ability to 'break the boundaries of work and home through their pervasiveness and 'always on' nature' (Ellis, 2004, p.8). While some profit-maximising small business owners welcomed this blurring of work and home boundaries, others did not, preferring to keep work and family spheres separate. Ellis cautions that in order to overcome existing negative preconceptions of AmI technologies, users must feel they control the devices and the data they produce, and be able to override them and cope with systems failure. In particular, AmI needs to be presented as "smarter" than existing ICTs and able to correct some of the problems associated with traditional forms of ICT support. In short, Ellis (2004, p. 9) asserts, "AmI needs to be associated with undoing some of the more problematic aspects of existing ICTs, to be accepted and not resisted as a more invasive, insidious and controlling form of what already exists."

Much of the promise of Ambient Intelligence (AmI) technologies rests upon connecting increasingly sophisticated and powerful sensors with existing computing facilities. McCullough (2001) identified the need to expand our thinking beyond the notion of filling environments with physical objects when considering Ambient Intelligence technologies. There is no longer such a thing as "empty space" when sensors and processing power combine to produce an environment that is "aware" of the locations, actions and information needs of humans. Clearly, the extent of existing Information and Communications Technology (ICT) infrastructure in an organisation will impact AmI technology implementations, providing either a "clean slate" from which to start or the opportunity to integrate new AmI capabilities with existing ICT systems and processes. In much the same way as the advent of mobile phones in China and India provided an opportunity for people unable to afford a landline, to access telephonic services, AmI technologies may prove a way for SME organisations to "leap frog" a stage of ICT implementation, and move directly to wireless and similar AmI technologies.

Many other applications of AmI technologies are appearing as technologists extend the concept into areas such as "wearable technology" (clothing that incorporates sensors and interface devices), more intuitive home space designs, shopping assistance and the creation of seamless interfaces between work, home and leisure activities. While many of these applications currently seem unrelated to improving business productivity, it is clear that the applications for business can only grow as the technologies become more sophisticated and less expensive. As Rao and Zimmerman (2005, p.3) state "there is a gap in the scholarly discussion addressing the business issues related to it, and the role of pervasive computing in driving business innovation". It is in this context that the following case studies of four small-to-medium (SME) manufacturers - three Australian and one German firm – have been undertaken. In each firm, critical process analysis was carried out to examine possible process weaknesses and existing ICT systems, and recommendations were made concerning a selection of AmI technologies with the potential to boost business performance.

AMBIENT INTELLIGENCE TECHNOLOGY IN MANUFACTURING

This section considers the applicability of several emerging AmI technologies to three SME manufacturers in New South Wales, Australia and compares the situation within these SMEs with one German SME manufacturer undertaking a similar technological adoption. In doing this the section also addresses questions about the preparedness of SMEs, particularly concerning their absorptive capacity limitations and how these may be overcome. Later sections also consider the potential impact of Ambient Technologies on the employees of the organisations studied.

AmI technology is much more than RFID inventory control systems. Wireless, multi-modal services and speech recognition systems have the potential to increase manufacturing flexibility by supporting dynamic reconfiguration of process and assembly lines, and improving human-machine interfaces to reduce process times (Maurtua et al., 2006). Also, maintenance and distribution processes may be improved by linking common mobile wireless devices, such as mobile phones, Personal Digital Assistants (PDAs) or even pagers to production alert systems (Stokic, Kirchhoff, & Sundmaeker, 2006).

Small and Medium Manufacturers in Australia

Organisations with between 20 and 199 workers employ 56% of Australia's workforce (Wiesner, McDonald, & Banham, 2007). The Australian Bureau of Statistics (ABS) defines a small business as employing less than 20 people, and a medium enterprise as employing between 20 and 200 employees (ABS, 2001). The most recent ABS figures available (2007) for Australia indicate that there are around 47,000 manufacturing firms employing between 1 and 20 people, around 10,000 employing between 20 and 200 people, and only 873 employing over 200 people. In turnover terms, around 29,000 manufacturing firms reported annual turnover between $500,000 and $10 million, while only 3,300 firms reported turnover of $10 million or above. It is clear that the bulk of manufacturing in Australia occurs in small-to medium firms. While SME firms employ the majority of manufacturing workers their expenditure on R&D notably lags behind that of large manufacturers. Within the manufacturing industry companies with more than 200 employees were responsible for 73% of total industry R&D expenditure, with only 27% being contributed by the SME sector (ABS, 2007).

However, in their exploration of the cost and impact of logistics technologies in US manufacturing firms Germain, Droge & Daugherty (1994) found that for manufacturing managers wanting to innovate with logistics technology, organisational size provides an advantage that transcends both the cost and nature of the technology. These authors confirmed the established view in 1994; that organisational size was positively correlated with technology adoption, as found in many previous studies. This link between manufacturing organisation size and increased ability to extract benefit from technological innovations may provide some explanation for the fact that while Australia's manufacturing output has quadrupled since the mid 1950s, the Australian Government Productivity Commission (2003) states that overall, it has not grown at the same rate as the service sector. The Productivity Commission also describes Australia's manufacturing sector as having "missed out on the productivity surge" of the mid 1990s while noting signs of improved manufacturing productivity in 2002 and 2003. The widespread availability of off-the-shelf ICT systems has probably meant that a great many more SMEs are adopting ICTs today than in 1994, however the limited resources of many such firms (both financial and human) almost certainly mean reduced awareness and limited capacity to exploit newer technologies commonly appearing in larger manufacturers. Given the significance of SMEs in Australian employment and the perceived need to increase manufacturing productivity, examination of the potential improvements available through the systemic application of AmI technology to SME manufacturers forms an important topic for research and government policy.

Absorptive Capacity

This chapter applies the concept of *absorptive capacity* to manufacturing SMEs. We argue that SMEs can benefit from AmI technologies, using specialised intermediary organisations to overcome the "absorptive capacity" limitations evident in many SME organisations. Cohen and Levinthal (1990) proposed that internal Research & Development activities serve two purposes: to generate innovations, and to provide the ability to absorb relevant knowledge appearing in the external environment. The absorptive capacity of a firm is comprised of these two categories of activity. Their foundational paper conceptualised absorptive capacity in the context of large U.S. manufacturers, as evidenced by their survey of identifiable "R&D lab managers" (Cohen & Levinthal, 1990, p. 142) and their discussion of "communication inefficiencies" between business units. But what of small- and medium-sized manufacturers? Does the notion of absorptive capacity have relevance to SMEs outside narrow, industry-segment specific technologies? If so, can external intermediaries assist SMEs to overcome absorptive capacity limitations regarding ambient technological innovations? The preparedness of SMEs to adopt and benefit from the mobile capabilities of AmI technologies is likely to be linked to their ability to overcome the limited absorptive capabilities associated with their small size.

However, it is possible for SMEs to successfully deploy AmI technologies. Clear evidence of the benefits produced by a range of AmI technologies have been observed by the researchers at Costa Logistics' distribution centre in Western Sydney. Costa Logistics specializes in the distribution of fresh fruit and vegetables, and achieves astounding levels of accuracy and throughput: staff average zero to three errors per million cartons picked in the warehouse, the daily inventory turn ratio is 100% and over A$2 billion of product is handled annually on behalf of their clients (Game-Lopata, 2008). The company was a SME when it started to use wrist-mounted bar-code scanners coupled with wireless communications. These successfully implemented AmI technologies are key factor that have enabled Costa Logistics rapid expansion to well over 200 employees in late 2009.

Innovation, Manufacturers, SMEs and Government Policy

Several previous studies (Cutler, 2008; Philips, 1997) have concluded that innovative Australian firms of all sizes (both manufacturing and service-based) tend to be more successful in terms of sales growth and market share than non-innovating firms. In addition, the impact of innovation is considered to be cumulative (Chapman, Toner, Sloan, Caddy, & Turpin, 2008) with some level of innovative behaviour or research and development being required to equip a firm to identify, assess and adopt technologies. The innovativeness and absorptive capacity of SMEs is a matter of concern for other nations besides Australia. For example, in its 2008 budget, the UK government signalled its intention to set a goal for innovative SMEs to win 30% of its £150 billion public procurement spending (Kable's Government Computing, 2008), equating to $98 billion (AUD) of incentives to encourage UK SMEs to innovate. While similar incentives are yet to appear in Australia, there are clear signs of government interest in the ability of SMEs to innovate (Department of Innovation Industry Science and Research, 2008).

There is a growing body of work in the innovation literature on the limited absorptive capacity of SMEs to identify relevant innovations, understand and appreciate possible applications, and finally adapt and implement innovation in their organisations (Beckett, 2008; Liao, Welsch, & Stoica, 2003; Muscio, 2008). Many points concerning "constraining factors" and "implementation challenges" support the notion that SMEs can experience organisational absorptive capacity limitations. Beckett (2008) identifies knowledge and resource constraints that impede the ability of SMEs to develop absorptive capacity, but also provides an example of how absorptive capacity is built when the outlays of time and money required match the SME's available resources.

While the benefits of AmI technologies are already accruing in large organisations (Angeles, 2005) if manufacturing SMEs are to benefit from AmI technologies, one challenge requiring attention will be that of their limited absorptive capacity for technological innovations. Our research considered the possibility of external intermediaries being used to facilitate SME manufacturers' assessment of the application of AmI technologies for process innovation, thus overcoming, at least partially, the problems of limited absorptive capacity within the partner SMEs.

METHODOLOGY AND DATA COLLECTION

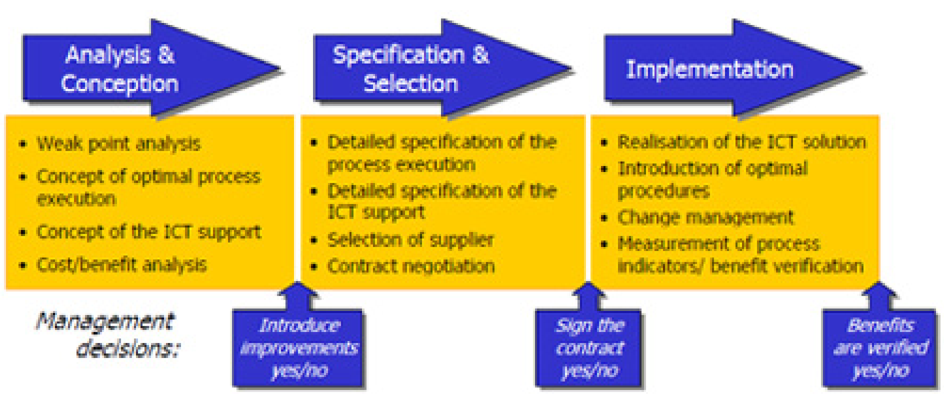
Case research has been used to review and compare the operations of the three NSW manufacturing SMEs, identify process weak points (in partnership with the SME executive managers) and suggest potential Ambient Intelligence Technologies to assist each organisation. The data has been gathered as part of an Australian government funded International Science Linkages Project, which in turn was part of a larger European-Australian project on Ambient Intelligence in manufacturing - Ambient Intelligence Technology for Systemic Innovation in Manufacturing SMEs (AmI-4-SMEs). The European component of the AmI-4-SMEs project involved six SMEs, three research partners and three Information and Communications Technology (ICT) providers located in Germany, Ireland, Spain and Poland. The Australian AmI-4-SMEs project consists of six SMEs, two research partners (University of Melbourne and University of Western Sydney) and two ICT providers. Six SMEs were selected from those responding to a request for expressions of interest in participating in the study. Using the Australian Bureau of Statistics metrics (2001), all are classified as small-to-medium sized manufacturers and all are privately owned.

The EU AmI-4-SMEs project aimed to design and develop a coordinated approach for process innovation using, ICT "building blocks" and a software platform to support the improvement of manufacturing processes in SMEs. These improvements were achieved by re-engineering processes and introducing appropriate ICT tools. The method used to analyse and re-engineer business processes is an extension of the COST-WORTH methodology (ATB Institute for Applied Systems Technology Bremen GmbH, 2004) and has three main phases: Analysis and Conception, Selection and Specification, and finally Implementation. Due to lower levels of government funding support, the Australian Ami-4-SMEs project performed only the Analysis and Conception, and the initial aspects of the Specification and Design phase but not the Implementation phase. The links between these phases are shown in Figure 1.

The Analysis and Conception phase produces an implementation plan for the proposed AmI solution. Analyses of each SME's business processes and bottlenecks form the majority of this phase which concludes with presentation of a business re-engineering recommendation and a Return on Investment Analysis (Kirchhoff, Stokic, & Sundmaeker, 2006).

One challenge of working with SMEs is to gather sufficient information without intruding to the extent that the organisation is adversely affected. On-site interviews and observation, questionnaires (these were developed as a part of the precursor

Figure 1. Representation of the Three Phase AmI-4-SME Methodology (*Source (Kirchhoff et al., 2006)*)



COSTWORTH project, see Nousala, Ifandoudas, Terziovski and Chapman, 2008), video recordings, a visit to a company (Costa Logistics) already using wireless, wearable and voice (Chang & Heng, 2006) technologies in its warehouse, and joint creation of process maps where they did not previously exist were used to collect data and minimise disruption to the SMEs. Interview and questionnaire data were used to select important, problematic processes for each SME. On-site observations and video recordings were analysed to create "as-is" maps of the process selected for improvement, and identify key limitations of each process. AmI technologies with potential to improve the selected business process were selected and the likely costs and benefits reviewed with the SME executive managers.

A strength (and simultaneous limitation) of this approach, is that the time spent at each SME site is not extended or intensive. However, given the objectives of the analysis and conception phase of the AmI-4-SMEs project, and the need to minimise disruption to the operation of the manufacturing businesses, the methods are appropriate.

RESULTS

Overview of the Three Australian SMEs

The three Sydney-based SMEs represent a wide range of existing ICT complexity and skill, from "craft-work" factories to highly sophisticated manufacturers of technology. Although the organisations were not intentionally recruited to represent a typology of low, moderate and highly integrated ICT installations and skill sets, they do display these characteristics and it is useful to consider how AmI technology capabilities may be added to each of these settings. Pseudonyms have been used for the SME companies. Ranked in order from low to higher technological capability they were: SwivelMould – a plastics manufacturer specialising in rotational moulding work; BottleTop – manufacturer of plastic packaging closures; and TechMakers – a contract electronics component manufacturer.

As described in the methodology section, due to lower funding levels, the Australian AmI-4-SMEs project did not fully incorporate the latter two phases of the project (see Figure 1), instead leaving it to the individual companies to complete these steps. The European project funding also included direct funding support for the SME partners, which was not provided under the Australian government support funds. However, the Analysis and Conception phase, and the initial aspects of the Specification & Selection phase, included a detailed initial assessment of each SME, identified possible AmI solutions and produced a rough implementation plan for the proposed AmI solution. Analyses of each SME's business processes and bottlenecks formed the majority of this phase, and these were conducted on-site at the three NSW SMEs.

Tables 1 and 2 present summary information and key issues related to technology adoption for each of the three SMEs from New South Wales, Australia.

Similarities and Differences between the Three Australian SME Manufacturers

Scalability issues related to specialised equipment appeared as a limitation shared by the three Australian SMEs. Production could not rapidly increase without the purchase of more machinery, and in the case of TechMakers, difficulty in arranging for the rapid importation of components also limited their growth.

Key differences observed in the three Australian SMEs relate to the proportion of standardised production and organisational culture. Eighty per cent of BottleTop's product volume results from standing contracts with its top twenty customers. Beyond the top twenty customers is what the Operations Manager refers to as a "long tail" of hundreds of small customers. In contrast, TechMakers survives

Table 1. Company summary

BottleTop	TechMakers	SwivelMould
Turnover: $30M **Employees** 97 **Profit**: moving from break-even to profitable trading in 2008/09. **Industry**: Manufacturing of packaging closures **Outlook**: Growing in a shrinking market by taking market share from competitors. BottleTop has set an aggressive revenue growth target of $50M by 2011/12.	**Turnover**: $7 M **Employees** 70 **Profit**: Trading profitably **Industry**: Contract Electronic Manufacturers. **Outlook**: Industry is shrinking, work is increasingly being sent off shore, TechMakers are diversifying by creating and selling the devices they created to assist in their contract electronic manufacturing activities.	**Turnover**: $20M **Employees**: 80 **Profit**: trading profitably, but recently burdened by a major bad debt. **Industry**: Rotational Moulding **Outlook**: Rapid growth during drought through manufacture of water tanks,

Table 2. Key issues related to technology adoption

BottleTop	TechMakers	SwivelMould
Staff The company managers view their employees as loyal, and based on a long association, almost like family. Some staff members resent the profits they believe the company owners are making and draw unfavourable comparisons to their hourly wage rate. A profit sharing scheme based upon reducing the scrap rate has been enthusiastically embraced by staff. BottleTop's labour efficiency has increased from 76 to 92% in the last year primarily attributable to their process focus as they implement Six Sigma Manufacturing techniques	No staff-related comments or concerns appeared in interviews or observations.	SwivelMould have a high employee turnover as the work is repetitive, takes place in a hot and noisy environment and does not pay high hourly rates. SwivelMould had difficulty recruiting and retaining employees while unemployment rates were low. New employees are only given formal training after they have completed a probationary period and if they appear likely to work at SwivelMould for some time.
Tracking of Products Product tracking is highly automated, integrated links between moulding machines and software packages running on PCs provide a real time view of activity.	The Quality Manager describes product tracking as "the black hole" because once a job commences no job status information is available until the end of the manufacturing run. This issue can be addressed by the organisation as TechMakers employ a full time in-house programmer. When time permits the programmer will interface product tracking with the MRP system.	Orders are received by fax and transcribed onto job sheets by hand. As jobs are completed on the shop floor the quantity, colour etc details are confirmed by the foreman writing on the same sheet. Any variations in process are also hand written on the sheet.
Process Improvements BottleTop is using six sigma methods to improve quality, reduce product variability and reduce waste. The company is focussing on improving each process prior to automating it. As the Operations Manager states "We want to have strong processes, we don't want rubbish processes just being done more quickly." When quality processes are in place BottleTop's focus will shift to automation and then to monitoring.	TechMakers processes are well developed and highly automated and with one exception, their work in progress system, discussed in the "Tracking of Products" column. A more difficult process to address concerns the performance of the distributor that acts as the exclusive Australian agent for the US manufacturer of components used by TechMakers.	SwivelMould has not mapped its processes. The process map built for one process as part of the AmI-4-SME process was the first time the company had used process mapping. Tradition and the knowledge of gang foremen are used to guide manufacturing processes.
Training Courses A mix of in-house and external training is used at BottleTop. External training is used to provide Six Sigma training.	No comments made in relation to training.	SwivelMould has developed a competency based training program in conjunction with an external training consultant. The objectives of the training are to increase productivity, quality and produce a change in the organisational culture.

continued on following page

Table 2. Continued

BottleTop	TechMakers	SwivelMould
Growth BottleTop managers are optimistic about their ability to grow revenue in a shrinking market. However, growing acceptance of closure systems make in countries with low labour rates may limits their growth.	TechMakers is and is the oldest and fourth largest contract electronic manufacturer (CEM) in Australia. The CEO is content to remain the fourth largest firm and claims there are advantages in not being the biggest player in the Australian market. The market for CEMs in Australia is shrinking as work moves off-shore. While the Quality Assurance Manager believes the company is committed to increasing revenue, the CEO (who is the owner of the company) stated privately that his objective is to improve the profitability of operations.	SwivelMould's ambitious expansion plans are on track to deliver the anticipated growth in revenue. However as rain water tanks provide 55-60% of annual revenue SwivelMould's plans are dependent to some extent upon the maintenance of government rebate policies, and continuation of Australia's ten year drought. During the course of the study the drought broke on the eastern seaboard, reducing demand for rainwater tanks. While newer entrants to the market have been unable to cope with the downturn in the tank business, SwivelMould's Managing Director is confident of its ability to continue its expansion due to investments made in new products and its pursuit of new markets.
Overseas Operations Sixty per cent of the company's revenue is derived from importation and distribution activities. Logistics is important and problematic for the company as products imported as finished goods create significant supply chain challenges. BottleTop is considering employing a logistics outsourcing firm to manage these challenges	TechMakers are pursuing export opportunities for the technologies they have invented in house. They do not intend to attempt to compete with CEMs located overseas.	A reduction in demand for water tanks has occurred with the breaking of the drought in two cities (Sydney and Brisbane). To some extent this has been offset by increases in exports; the quantity of goods exported to the USA and China has trebled in the last year.
Documentation & Processes Highly integrated with and automated by ICT systems.	Highly integrated with and automated by ICT systems.	Rudimentary documentation, handwritten and physically carried between office and factory floor. Little integration with ICT systems.

in a shrinking market due to its flexibility, rapid prototyping turn-around and high quality output. The CEO and Quality Assurance manager concur that "every day is different." SwivelMould's operations sit between the other two SMEs. SwivelMould offers a full service from product concept through design to manufacture. As each type of product requires a specialised mould, batch runs are used in production. While safety equipment is in use at TechMakers and BottleTop, SwivelMould employees resist using standard safety equipment such as hard hats and hearing protection. Only foremen wear high visibility vests. The Managing Director describes these actions as symptomatic of the "70s culture" that he is trying to change.

AmI Technology Recommendations

The Australian SMEs

The wide range of technical skill levels in the three Australian SME manufacturers results in very few similarities in their technological capacities. The staff in two of the SMEs are happy with the opportunity to enhance their skills and comfortable with process change. In all the SMEs the arrival of technology that is industry-specific is seen as a form of "natural progression" from previous machines. In one SME (SwivelMould) the CEO is experienced difficulty in moving staff away from established skills and procedures.

The recommendations for the SMEs were tailored to the existing environment of each organisation in terms of industry segment and their legacy ICT systems. In the case of SwivelMould, recommendations were made for AmI technologies that either link together existing "islands" of ICT equipment or provide quickly implemented, stand-alone solutions to environmental issues. For example, a recommendation to implement an integrated manufacturing system, taking advantage of the flexibility of wireless communications was made. As Kirchoff, Stockic and Sundmaeker (2006) assert, if insufficient or poorly integrated ICT infrastructure exists the first step towards obtaining the benefits of AmI technologies is to introduce ICT systems to support general manufacturing processes. Given the size of SwivelMould's operations, it is likely that a standardised manufacturing package working on a PC/Windows, Unix or Linux operating environment will address this need. To provide a small example of the benefits available, using a secure on-line order entry system will avoid the need to manually transfer order details faxed to SwivelMould by their distributors onto worksheets.

The second organisation, TechMakers presents a challenge as their business is manufacturing electronic devices and sub-assemblies, effectively acting as an outsourced technology design and manufacture facility for their clients. TechMakers are well aware of AmI technologies and would have already incorporated them into their operations if they had identified potential applications. Furthermore, their key business issues concern their contracting market and their dependency upon an intermediary distributor to order components from the USA. For TechMakers, the benefits available from AmI technology may exist in opportunities to design, manufacture or modify AmI systems for other companies.

At BottleTop an alert system based on AmI technologies in the form of wireless communications has the potential to improve productivity by decreasing the number and duration of production stoppages caused by machine parameters moving out of set tolerances, and release personnel from repetitive inspection tasks for higher value, and more rewarding and interesting work. The following section compares the detailed findings from one Australian SME (BottleTop) with those from a German SME participating in the European AmI-4-SMEs project. Despite differences in industry and location, both these SMEs have very similar opportunities to address production issues using AmI Technologies, pointing to the potential for standardised AmI based solutions to improve SME manufacturing operations.

Comparison between European and Australian SME Manufacturers

German SME (Truckbody GmbH)

Truckbody GmbH claims market leadership for EU manufacture of truck swap bodies (steel framed transport containers, and the legs on which they stand while waiting transfer from truck to truck, or truck to rail) primarily intended for the EU domestic market. A key competence is the manufacture and powder coating of large structures, up to 15m long, such as bus frames. The company employs 330 people, which places it in the EU classification of SME organisations, in contrast to Australia where a SME is classified as an organisation with between 20 and 200 employees. Truckbody's production system is characterized by strong interdependencies between different task groups; a delay in one step impacts many other groups further down the production line.

To reduce production delays the EU Ami-4-SMEs project research and technology partners identified a need for automatic production alerts that interfaced to the company's planning system. The ATB Institute for Applied Systems Technology, based in Bremen, Germany is the project leader of the EU AmI-4-SMEs project, and is currently finalising the implementation of a rule engine and user interfaces on mobile devices. When problems occur in Truckbody's production, employees who need to know about the disruption, such as the shift supervisor and the person with the skill to solve the problem, receive an automatically generated alert message. The alerts are based on user profiles (e.g. manager, foreman), the current location of the user (e.g. meeting, office, home) and the severity of the situation (e.g. deviation threshold, breakdown, loss). Use of a multi-modal user interface, (specifically a wireless message sent to a mobile phone or PDA) leverages the capability of AmI technologies to provide timely alerts that are "pushed" to relevant employees regardless of their location. In this manner, the AmI technology provides immediate and mobile access to production information, warns of delays to the production line, and so supports reallocation of work and staff. Prototypes have been developed as part of the Selection and Specification phase of the study.

Australian SME (BottleTop Pty Ltd)

BottleTop produces a very different product from that of Truckbody GmbH. BottleTop manufactures specialty packaging, with particular strengths in the personal care, pharmaceutical, health foods, chemical, cleaning, food, beverage and cosmetics markets. Operating for sixty years from its single Sydney manufacturing site, it has built a strong sense of loyalty among its 97 employees, and has extensive links to international fastening manufacturers.

Although plastic manufacturing accounts for around 7% of all Australian manufacturing activity, the industry is quite mature (McCaffrey, 2006), and is shrinking at around 4% per year, mainly due to increased purchases from foreign injection moulding companies. BottleTop is growing in this shrinking market, winning market share from its competitors by focussing on quality, service, technology and relationships within and outside the organisation. The company plans to more than double its revenue by 2011/12. While the revenue goal is ambitious, BottleTop's revenue grew by 12% in 2006 even after allowing for a 5% reduction in revenue from its existing customer base due to some customers moving their business to off-shore suppliers.

Discussions with BottleTop's Production Management team identified the following AmI technology scenario as an attractive business concept: BottleTop's moulding and assembly machines have in-built Programmable Logic Controllers (PLCs) which can monitor the six key variables that control the formation of the plastic closure. If a software program collects and monitors the PLC data, when any of these six parameters move outside pre-set limits an SMS alert to a mobile phone, or pager message could be automatically generated and sent to on-site maintenance personnel. This provides several potential business benefits, including minimization of machine downtime, reduction of defective, scrapped product and reduced need for visual inspection. Currently all the plastic fasteners are inspected by a human operator as they leave the machine.

Previous attempts to use computers coupled to cameras to replace human visual inspection of parts leaving the injection moulding machines were not successful due to the camera's inability to cope with reflective foil routinely used in BottleTop's products. It is important to note that the company's HR practices are likely to support the introduction of the proposed AmI solution. A bonus scheme rewarding operators for reducing the amount of defective caps produced from each machine has been enthusiastically embraced; operators have been heard to comment, "That's my money on the floor" when the speed of the machine is set too fast and fasteners overshoot the hopper.

The preceding comparison demonstrates that despite operating in unrelated industries in different countries, some SME manufacturing processes have sufficient commonality to permit the development of generic AmI solutions. Furthermore, the appearance of the same requirement in different manufacturing contexts shows that AmI technologies have the potential to be "general purpose" production enablers in diverse SME manufacturing settings. This in turn suggests the possibility that affordable "turn-key" AmI solutions may become available from technology providers. The next section considers the possibility of third party technology providers tailoring generic AmI solutions to the specific requirements of each SME, thus overcoming the absorptive capacity limitations inherent in SMEs.

DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

In Australian manufacturing SMEs, there is a very low likelihood of in-house R&D being used to build absorptive capacity to investigate AmI technologies. SMEs prefer to buy new technology when it is already embedded in an industry specific product rather than master the details of the underlying innovation (Oyelaran-Oyeyinka & Lal, 2006). Instead, we propose that SMEs are more likely to use industry or informal networks to become aware of potentially useful innovations, and then "buy" the innovation embedded in capital equipment or consulting services as a means to 'recognise the value of new information, assimilate it, and apply it to commercial ends' (Cohen & Levinthal, 1990, p.128).

However, Cohen and Levinthal (1990) question the effectiveness of "buying" absorptive capacity in the form of consulting services or through acquisitions when the knowledge is to be integrated with existing business systems. They state "To integrate certain classes of complex and sophisticated technological knowledge successfully into the firm's activities, the firm requires an existing internal staff of technologists and scientists who are both competent in their fields and are familiar with the firm's idiosyncratic needs, organizational procedures, routines, complementary capabilities, and extramural relationships" (Cohen & Levinthal, 1990, p. 135).

Out-sourcing of deep absorptive capacity to equipment and software vendors able to provide "turn-key" solutions that match industry requirements seems to be a way for manufacturing SMEs to gain the commercial benefits of AmI technologies despite the resource and time constraints that prevent them building absorptive capacity in any area other than their core business competence. Similar requirements appear in two very different SMEs on two continents. The potential for the same AmI technology solution components to address these requirements, albeit tailored to the specifics of equipment in use at each site, suggests that SMEs can benefit from AmI technologies by using specialised intermediary organisations to provide the "absorptive capacity" on their behalf. This finding points to potential links between absorptive capacity and "make vs. buy" decision-making, and to "broad" or "deep" versions of absorptive capacity (Henard & McFadyen, 2006) as avenues for future research. In addition, an opportunity exists to track the spread of AmI technologies in SME Australian manufacturers and in doing so contribute to the diffusion of innovation literature.

Additionally, AmI implementation challenges for Australian SME manufacturers extend beyond the boundaries of their own organisations. Large ICT manufacturers use a channel marketing approach to sell their products to the SME market segment. The channels may include retail and direct sales forces, but frequently hardware is "bundled" with service and software offerings from business partners, specialising in a particular industry segment, such as manufacturing. While intermediary business partners may supply specialised knowledge and generic AmI solutions to compensate for limited SME absorptive capacity, the organisations that partner with large ICT providers are often SMEs themselves. The ability and willingness of these business partners to gain AmI skills may in turn be a limiting factor in the adoption of AmI technologies by Australian Manufacturing SMEs. Absorptive capacity limitations of SME organisations can potentially affect uptake of AmI technologies at two points: within the manufacturing SME and within the SME technology partner. Low levels of in-house AmI skills and heavy level reliance on SME Australian technology providers suggest there may be an argument for the provision of government subsidies to encourage the adoption of AmI technologies in Australian manufacturing. A precedent exists in that subsidies have been provided for the purchase of RFID scanners for NSW meat producers (NSW Farmers Association, 2007). Without some form of government encouragement, the task of integrating AmI systems with existing ICT investments and the concomitant diversion from core manufacturing activities, may be enough to prevent the adoption of AmI technologies and, therefore, achievement of the elusive "productivity surge" in Australian manufacturing SMEs.

CONCLUSION

Advances in the development of ICT industry standards, and the proliferation of software and support for the Windows/Intel platform since Cohen and Levinthal's 1990 paper have brought technology to SMEs without the need for bespoke development. Furthermore, Cohen and Levinthal appear to assume that investments in absorptive capacity only exist in the form of R&D spending, rather than networking with other organisations to use "connect and develop" models typical of Open Innovation (Chesbrough, Vanhaverbeke, & West, 2006). In contrast, the results from the EU and Australian AmI-4-SME projects (ATB Institute for Applied Systems Technology Bremen GmbH, 2008) suggest that SMEs can use "external research sub-units" in the form of experiences reported by members of their industry network and trade associations, and solutions proposed by research and technology providers, to offset internal absorptive capacity limitations.

ACKNOWLEDGMENT

The authors would like to thank the SME participants in the Australian AmI-4-SME project for providing access to, and information about their organisations. We would also like to thank our colleagues in Australia, Assoc. Prof. Mile Terziovski and Richard Ferrers at the University of Melbourne, Dr David Low at the University of Western Sydney and our colleagues at the ATB Institute for Applied Systems Technology; Bremen, Germany for their helpful advice and the opportunity to compare Australian and EU based SMEs. The authors would also like to acknowledge the support of an International Science Linkages grant from the Australian Department of DIISR (Project Number CG110181), the University of Western Sydney and the University of Melbourne.

REFERENCES

Angeles, R. (2005). RFID Technologies: Supply-Chain Applications and Implementation Issues. *Information Systems Management*, *22*(1), 51–65. doi:10.1201/1078/44912.22.1.20051201/85739.7

ATB Institute for Applied Systems Technology Bremen GmbH. (2004). COST-WORTH Methodology Independent Reference Scheme [Electronic Version]. Retrieved April 22, 2008 4:25 pm from ATB Bremen GmbH.

ATB Institute for Applied Systems Technology Bremen GmbH. (2008). AMI-4-SME Platform. from http://ami4sme.org/results/platform.php

Australian Bureau of Statistics. (2001). *1321.0 Small Business in Australia*. Retrieved April 21, 11am, 2008, from http://www.abs.gov.au/Ausstats/abs@.nsf/0/97452F3932F44031CA256C5B000 27F19?Open

Australian Bureau of Statistics. (2007). *8104.0 - Research and Experimental Development, Businesses, Australia, 2005-06*. Retrieved September 3, 2:50pm, 2008, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/8104.0Main%20Features32005-06?opendocument&tabname=Summary&prodno=8104.0&issue =2005-06&num =&view=

Australian Bureau of Statistics. (2009). 8129.0 - Business Use of Information Technology, 2007-08 (Publication. Retrieved September 17, 2009: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/29F02AF6A0F6C9A3CA2576170018FB56?opendocument

Australian Productivity Commission. (2003). Trends in Australian Manufacturing. Retrieved April 21, 11:46 am, 2008, from http://www.pc.gov.au/research/commissionresearch/tiam/keypoints

Australian Productivity Commission. (2004). *ICT Use and Productivity: A Synthesis from Studies of Australian Firms*. Retrieved. From http://www.pc.gov.au/research/commissionresearch/ictuse

Beckett, R. C. (2008). Utilizing and Adaptation of the Absorptive Capacity Concept in a Virtual Enterprise Context. *International Journal of Production Research*, *46*(5), 1243–1252. doi:10.1080/00207540701224327

Brown, S., & Bessant, J. (2003). The Manufacturing Strategy-Capabilities Links in Mass Customisation and Agile Manufacturing - an Exploratory Study. *International Journal of Operations & Production Management*, *23*(7), 707–730. doi:10.1108/01443570310481522

Campos, A., Pina, P., & Neves-Silva, R. (2006). Supporting Distributed Collaborative Work in Manufacturing Industry. IEEE Xplore. Retrieved Feb 22, 2010

Caprio, D. W. J. (2005). *Radio-Frequency Identification (RFID): Panorama of RFID Current Applications and Potential Economic Benefits*. Paper presented at the Committee for Information, Computer and Communication Policy (ICCP) of the Organization for Economic Co-operation and Development (OECD). Retrieved Viewed April 14, 2008, from http://www.oecd.org/dataoecd/60/8/3546 5566.pdf

Chang, S. E., & Heng, M. S. H. (2006). An Empirical Study on Voice-Enabled Web Applications. Pervasive Computing (July-September 2006), 76 - 81.

Chapman, R., Toner, P., Sloan, T., Caddy, I., & Turpin, T. (2008). *Bridging the Barriers: A Study of Innovation in the NSW Manufacturing Sector*. Sydney: University of Western Sydney.

Chesbrough, H. W., Vanhaverbeke, W., & West, J. (2006). *Open Innovation: Researching a New Paradigm*. Oxford: Oxford University Press.

Chong, S., & Pervan, G. (2007). Factors Influencing the Extent of Deployment of Electronic Commerce for Small-and Medium-Sized Enterprises. *Journal of Electronic Commerce in Organizations*, *5*(1), 1–29.

Cochran, P. L., Tatikonda, M. V., & Magid, J. M. (2007). Radio Frequency Identification and the Ethics of Privacy. *Organizational Dynamics*, *36*(2), 217–229. doi:10.1016/j.orgdyn.2007.03.008

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152. doi:10.2307/2393553

Cutler, T. (2008). *Venturous Australia: Building Strength in Innovation; Report of the Review of the National Innovation System in Australia*. Melbourne: Cutler and Company.

Department of Innovation Industry Science and Research. (2008, Feb 19, 2008). Small Business Surveys - Finance and Banking, Innovation, and International Activity Retrieved June 3, 2008, from http://www.innovation.gov.au/Section/SmallBusiness/Pages/SmallBusinessSurveysFinanceandBankingInnovationandInternationalActivity.aspx

Ellis, R. M. (2004). The Challenges of Work/Home Boundaries and User Perceptions for Ambient Intelligence [Electronic Version]. *Chimera Working Paper, 2004-14. Colchester: University of Essex*. Retrieved July 1, 2008, from http://www.essex.ac.uk/chimera/content/pubs/wps/CWP-2004-14-RE-e-Challenges.pdf

Game-Lopata, A. (2008). Exclusive: Secret Sauce [Electronic Version]. *Logistics*. Retrieved September 18, 2009, from http://www.logisticsmagazine.com.au/Article/Exclusive-Secret-Sauce/172799.aspx

Germain, R., Droge, C., & Daugherty, P. J. (1994). A Cost and Impact Typology of Logistics Technology and the Effect of its Adoption on Organizational Practice. *Journal of Business Logistics*, *15*(2), 227–248.

Gunasekaran, A., & Yusuf, Y. Y. (2002). Agile Manufacturing: a Taxonomy of Strategic and Technological Imperatives. *International Journal of Production Research*, *40*(6), 1357–1385. doi:10.1080/00207540110118370

Henard, D. H., & McFadyen, M. A. (2006). R & D Knowledge *is* Power. *Research Technology Management*, *49*(3), 41–47.

Kable's Government Computing. (2008). Chancellor sets 30% Target for SME Procurement [Electronic Version]. *KableNET*. Retrieved April 21, 12:10pm from www.kablenet.com/kd.nsf/FrontpageRSS/23B6ED42820897968025740A004A2E25!Open Document

Kinder, T. (2002). Emerging E-commerce Business Models: an Analysis of Case Studies from West Lothian, Scotland. *European Journal of Innovation Management*, *5*(3), 130–151. doi:10.1108/14601060210436718

Kirchhoff, U., Stokic, D., & Sundmaeker, H. (2006). *AmI Technologies Based Business Improvement in Manufacturing SMEs*. Paper presented at the eChallenges e-2006. from http://www.ami4sme.org/.

Kopacsi, S., Kovacs, G., Anufriev, A., & Michelini, R. (2007). Ambient Intelligence as Enabling Technology for Modern Business Paradigms. *Robotics and Computer-integrated Manufacturing*, *23*(2), 242–256. doi:10.1016/j.rcim.2006.01.002

Kuehnle, H. (2007). Post Mass Production Paradigm (PMPP) Trajectories. *Journal of Manufacturing Technology Management*, *18*(8), 1022–1037. doi:10.1108/17410380710828316

Li, X., Feng, L., Zhou, L., & Shi, Y. (2009). Learning in an Ambient Intelligent World: Enabling Technologies and Practices. *IEEE Transactions on Knowledge and Data Engineering*, *21*(6), 910–924. doi:10.1109/TKDE.2008.143

Liao, J., Welsch, H., & Stoica, M. (2003). Organizational Absorptive Capacity and Responsiveness: An Empirical Investigation of Growth-Oriented SMEs *Entrepreneurship*. *Theory into Practice*, *28*(1), 63–85.

Maurtua, I., Perez, M. A., Susperregi, L., Tubio, C., & Ibarguren, A. (2006). *Ambient Intelligence in Manufacturing*. Paper presented at the Intelligent Production Machines and Systems, 2nd I*PROMS Virtual Conference.

McCaffrey, J. (2006). Version]. Retrieved Dec 17, 2007 from http://commercecan.ic.gc.ca/scdt/bizmap/interface2.nsf/vDownload/ISA_5111/$file/X_4862570.PDF

McCullough, M. (2001). On Typologies of Situated Interaction. *Human-Computer Interaction*, *16*, 337–349. doi:10.1207/S15327051HCI16234_14

Muscio, A. (2008). The Impact of Absorptive Capacity on SMEs' Collaboration. *Economics of Innovation and New Technology*, *16*(8), 653–668. doi:10.1080/10438590600983994

Nousala, S., Ifandoudas, P., Terziovski, M., & Chapman, R. L. (2008). Process Improvement and ICTs In Australian SMEs: A Selection and Implementation Framework. *Production Planning and Control*, *19*(8), 735–753. doi:10.1080/09537280802476169

NSW Farmers Association. (2007). National Livestock Identification System (Cattle) 013.07i NLIS Cattle [Electronic Version]. Retrieved April 10, 2008 from http://www.nswfarmers.org.au/data/assets/pdf_file/0012/3072/FS_NLIS_Cattle_0207.pdf

Oyelaran-Oyeyinka, O., & Lal, K. (2006). *SMEs and New Technologies; Learning E-Business and Development*. Basingstoke, UK: Palgrave McMillan. doi:10.1057/9780230625457

Philips, R. (1997). *Innovation and Firm Performance in Australian Manufacturing*. Industry Commission Staff Research Paper. Canberra, AGPS

Rao, B., & Zimmermann, H.-D. (2005). Pervasive Computing and Ambient Intelligence: Preface [Electronic Version]. *Electronic Markets*, 15, p. 3. Retrieved June 3, 2008 from http://www.informaworld.com/smpp/ftinterface~content=a713735017~fulltext= 713240928.

Stokic, D., Kirchhoff, U., & Sundmaeker, H. (2006). *Ambient Intelligence in Manufacturing Industry: Control System Point of View*. Paper presented at the IACTED Conference on Control and Applications 2006 conference

Vasilakos, A. V. (2008). Ambient Intelligence. *Information Sciences*, *178*(3), 585–587. doi:10.1016/j.ins.2007.08.016

Weber, W. (2003). Ambient Intelligence - Industrial Research on a Visionary Concept. IEEE Xplore Retrieved Feb 22, 2010

Wiesner, R., McDonald, J., & Banham, H. C. (2007). Australian Small and Medium Sized Enterprises (SMEs): A study of high performance management practices. *Journal of Management & Organization*, *13*(3), 227–248. doi:10.5172/jmo.2007.13.3.227

Zuboff, S. (1988). *In the Age of the Smart Machine: the Future of Work and Power*. Oxford Heinemann Professional.

KEY TERMS AND DEFINITIONS

Absorptive Capacity: The absorptive capacity of a firm is comprised of its ability to generate innovations, and absorb relevant knowledge appearing in the external environment.

Ambient Intelligence Technologies (AmI): Technologies that combine to create an environment that is are sensitive and responsive to the presence of people

Open Innovation: Organisations using ideas and capabilities originating outside their boundaries in order to increase the rate with which innovation occurs and decrease innovation costs. Open innovation also includes an organisation selling innovative ideas it has generated but cannot use in its business.

Pervasive and Ubiquitous Computing: Terms in use in the USA to refer to the same technologies as those named Ambient Intelligence Technologies in Europe

Radio Frequency Identification Device (RFID): Data collection devices consisting of electronic tags for storing unique identifying data.

Small Medium Enterprise SME: Small to Medium Enterprise. This measure of a company's size is generally based upon employee numbers, and varies across countries.

Technology Adoption: A process that begins with awareness of a specific type of technology or device, and progresses through stages ending in use or rejection of the technology.

This work was previously published in Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts, edited by Maria Manuela Cruz-Cunha and Fernando Moreira, pp. 65-82, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 57
Teaching Technology Computer Aided Design (TCAD) Online

Chinmay K Maiti
Indian Institute of Technology, India

Ananda Maiti
Indian Institute of Technology, India

ABSTRACT

Since Technology Computer Aided Design (TCAD) is an important component of modern semiconductor manufacturing, a new framework is needed for microelectronics education. An integrated measurement-based microelectronics and VLSI engineering laboratory with simulation-based technology CAD laboratory is described. An Internet-based laboratory management system for monitoring and control of a real-time measurement system interfaced via a dedicated local computer is discussed. The management system allows the remote students to conduct remote experiments, perform monitoring and control of the experimental setup, and collect data from the experiment through the network link as if the student is physically in a conventional laboratory. The management system is also capable of evaluating of a student's performance and grading laboratory courses that involve preliminary quiz and viva-voce examinations, checking of experimental data and submitted online laboratory reports. The proposed online TCAD teaching methodology will provide an opportunity for expanding microelectronics education.

INTRODUCTION

The field of microelectronics technology is recognized as a driving force for the Information Age. Micro- and nanoelectronics device and circuit design and fabrication are specialized fields in electrical engineering. The main goal of undergraduate and/or postgraduate level microelectronics teaching is to produce high-quality engineers who are able to make contributions in the context of the rapid change that characterizes integrated circuit (IC) fabrication. For microelectronics courses, laboratory should include a clean room infrastructure, semiconductor equipment operation procedures, process and metrology, device testing, and process integration and manufacturing learning as hands-on fabrication as well as characterization of devices that enhance the educational

DOI: 10.4018/978-1-4666-1945-6.ch057

experience. However, due to the high cost of a microelectronic fabrication laboratory, teaching microelectronic circuit fabrication is very much driven by the availability of resources at the institution providing such courses and is primarily taught at universities where an actual fabrication facility is available and, currently, it is mostly taught via demonstration mode. Microelectronics engineering education is in transition. New thought is being given to topics such as what constitutes microelectronics process design fundamentals, how to shrink the gap between industrial and academic perspectives on process design, and how to help students gain more experience and knowledge.

Currently, in most final year undergraduate and virtually all master's level programs, there are courses on device physics and processing technology (usually as one single course) based on standard text books on MOS and bipolar device physics that often do not include a laboratory component. Integrated circuit fabrication courses are offered as an elective in some Electrical and/or Electronic Engineering programs that cover fabrication theory of integrated circuits and process modeling. The introduction of process and device simulations in undergraduate teaching is also considered a difficult task. This is mainly due to the complicated user interaction with most of the available process and device simulators; usually the input information is prepared in the form of files written in a specific input language for each simulator. In general, professional Technology CAD (TCAD) simulation tools are difficult to use and are considerably more complex. The users need dedicated training sessions to successfully use the tools.

Also, during the last three decades, a new generation of semiconductor processing involving new material systems such as strained-Si and Silicon-Germanium (SiGe) have appeared, and integration of Group III-V compound semiconductors with Si technology is evolving (Maiti, Chattopadhyay, & Bera, 2007). With these advancements in semiconductor manufacturing, it is becoming difficult for the VLSI designers to optimize circuit design without considering the effects of advanced ULSI/GSI integration processes. The International Technology Roadmap for Semiconductors (ITRS, 2007) predicts the use of TCAD may provide as much as 40% reduction in technology development costs. TCAD has grown in both sophistication and maturity and is now an essential engineering tool for new technology development in industrial environments. Recent industry trends have given rise to major development opportunities for TCAD, and the virtual wafer fabrication has now become an integral part of semiconductor fabrication.

THE NEED

Semiconductor device theory and IC processing courses are becoming more important in electrical engineering curricula due to the fast changing semiconductor technologies and the challenges faced by the semiconductor industry. Process/device simulation tools are being introduced to students taking graduate courses in device physics and SPICE for circuit simulation. Semiconductor CAD originated in the early 1960s with efforts to understand and optimize bipolar transistors (Dutton, & Yu, 1993). TCAD is now a proven approach for developing process technologies, and a comprehensive set of TCAD tools is available from universities and from TCAD vendors. Advanced TCAD tools are capable of modeling the entire semiconductor manufacturing process. Also, the most challenging issue in the context of IC design and manufacturing now is product yield because the main causes for yield loss have changed over the years. Technology development in the semiconductor industry is a very complex task and requires a deep understanding of different physical and mathematical techniques. Next generation semiconductor engineers will need to be competent in the use of advanced process and device simulation tools as part of their everyday practice, as TCAD simulation allows a more transparent vision of state-of-the-art IC fabrication. The renewed focus on chip interconnects is moving TCAD modelling beyond active devices and further into the higher levels of the EDA tool hierarchy.

It is imperative that microelectronics education introduce industry-standard computer aided design tools to enhance traditional teaching methods and to provide students with the necessary CAD skills as required for semiconductor industry. TCAD software requires a considerable investment both in development of new capabilities and training in the use. Several attempts have been made to integrate professional TCAD simulation tools in semiconductor device courses (Kenrow, 2004). The students are also encouraged to take an IC fabrication laboratory along with the course. However, available TCAD tools are dispersed in nature and problems are encountered in locating, acquiring, installing, and learning to operate a proper simulation tool. The other difficulty is to provide the students with access to TCAD simulation tools through user-friendly computing platforms, which now tend to be personal computers to support the process/device simulation and modeling activities.

Availability of high-speed internet and advances in Information and Communication Technology (ICT) has allowed the enhancement of traditional learning methods. Advances in computer technology have made it feasible to provide engineering students with computer support for learning, especially in technical education. Nowadays, many universities have developed e-learning materials for theory courses and online laboratory facilities allowing students to perform laboratory experiments by simulation or with remote equipments (Maiti, 2010, Fjeldly, & Shur, 2003). Internet accessible device simulation laboratories for semiconductor education, which is a virtual laboratory accessible via standard World Wide Web (www) browsers, have also been reported and provide students using different computing environments access to a broad range of simulation tools (for example, http://www.ecn.purdue.edu/labs/punch). Users can define simulations, run them, and view the printed and graphical output.

To take full advantage of the predictive power of TCAD to reduce cost and time in semiconductor process development, introduction of "Technology CAD" as a subject in final year undergraduate and graduate level microelectronics courses was introduced in 1998 at IIT Kharagpur. In this chapter, we shall present a summary of our experiences gained in teaching TCAD during the last ten years. Emphasis is given to the critical role of the integrated online TCAD laboratory toward VLSI design along with process and device modelling. The chapter will address the issues and problems related to the intranet and/or internet to be used as the network for TCAD simulation laboratory for the remote users. Technology CAD is a one-semester, three credit (35-40 lecture hours and 30 laboratory hours) course for graduate students enrolled in Microelectronics and VLSI Design at IIT Kharagpur. The curricula are designed in such a way that the students are exposed to design procedures by participating in device and process designs via simulation. The course covers various topics via lectures, homework, laboratory assignments, and design projects that give the students the opportunity to gain extensive knowledge in semiconductor device/circuit design and fabrication. A brief review of the TCAD applications in semiconductor manufacturing is made first. The TCAD course structure and the objectives of the laboratory session are presented next. Use of an integrated measurement-based online Microelectronics and VLSI Engineering Laboratory and simulation-based Technology CAD Laboratory for expanding microelectronics education is described in detail.

What is Technology CAD?

Simulation of the fabrication steps of microelectronic devices and integrated circuits is a technique that is used extensively in silicon IC manufacturing to speed product development and reduce the development costs. Technology computer aided design is the term used to describe a broad range of modeling and analysis activities that consist of detailed simulation of IC fabrication processes, device electrical performance and extraction of device parameters for equivalent circuit models. Semiconductor manufacturing companies use TCAD for monitoring, analyzing and optimizing the IC process flows. The demand for engineers with experience in TCAD is increasing. Possessing a working knowledge or proficiency with TCAD is rapidly becoming a highly marketable skill for process engineers. Introduction of TCAD tools in device physics courses is becoming a necessity to provide the students with critical skills in the microelectronics area.

Two major components of TCAD are (a) Process simulation and (b) Device simulation. Process simulation is modeling of semiconductor manufacturing processes. Process simulation allows one to investigate the possible process steps prior to actual fabrication. Process simulation includes basic unit steps such as oxidation, implantation, diffusion, etching, growth, and deposition. Process simulation can also highlight issues that could potentially lead to circuit failure. Simulations allow verification of process design rules with the goal of design for manufacturability. Device simulation is for modeling the semiconductor devices. Device simulation produces the dc and ac characteristics of devices and is used to determine transistor models. SPICE model parameters are extracted from the measured electrical characteristics and the model parameters can be used in circuit simulators such as SPICE, which is often used as a key tool in Electronic Computer Aided Design (ECAD) to design VLSI circuits.

As TCAD helps in the design of process technology, it is often used in research and development of new processes and devices using computers as a substitute for the semiconductor processing laboratory. In process modeling, a systematic design of experiments (DoE) run is performed. DoE experiments can be systematically set up with control over process parameters and arbitrary choice of device performance characteristics. Most TCAD simulation tools provide detailed, colorful plots of device profiles and processing steps, giving the student the ability to see what is happening inside the device and to understand how the details of fabrication affect device properties. TCAD also offers greater physical insight into device operation; for example, one can visualize parameters that cannot be measured, such as current or electric field distributions inside a device. A costly fabrication laboratory is not needed for the purpose of demonstrating the fabrication steps as a computer is used as a substitute for a semiconductor processing laboratory. Students can also perform TCAD simulations beyond normal laboratory hours and at their own pace. Currently available professional TCAD tools are a very powerful and economical means for teaching microelectronics technology.

The area of Design for Manufacturing (DFM) is also becoming very important. The idea of DFM is to shorten the design cycle, to reduce manufacturing costs, and to improve yield, resulting in faster time-to-market and increased profitability. TCAD also addresses the issues of process and device variability in manufacturing to mitigate the rising costs of semiconductor process and product development as each technology node is driving towards higher performance and increased complexity. TCAD may be used as a powerful tool to identify root causes for yield loss and are used to study device sensitivities on process variations. Currently a major trend in the industry is to apply TCAD tools far beyond the integration phase into manufacturing and yield optimization. Linking of process parameter variations with the electrical parameters of a device through Process Compact Model (PCM) is in use. Process Compact Models are necessary to enable efficient analysis of complex and multivariate process-device relationships, with applications to enhancing process manufacturability and process control. PCM may be used to capture the nonlinear behavior and multi-parameter interactions of manufacturing processes. This is driven by the fact that product yield loss nowadays is dominated by systematic defects coming from lithography and IC design as well as by random process variations. For an extended TCAD in process modeling, generally a systematic design of experiments run is performed. DoE experiments can be systematically set up to study the control over process parameters and arbitrary choice of device performance characteristics. SPICE process compact models (SPCMs) can be considered as an extension of PCMs applied to SPICE parameters. By combining calibrated TCAD simulations with global SPICE extraction strategy, it is possible to create self consistent process-dependent compact SPICE models with process parameter variations as explicit variables.

The simulation experiments in TCAD have the advantage that every process condition can be accurately controlled and that arbitrary product performance characteristics can be determined. This is opposed to real experiments, where the control of process steps may be difficult and subject to uncontrollable drift or variation in the equipment and where the limitations of metrology can make it difficult, expensive or impossible to make both nondestructive and destructive measurements. Visualization is one of the big advantages of TCAD tools, as during process simulation, evolution of actual cross-sections of the structure can be observed.

SPICE Parameter Extraction

The success of a VLSI circuit design depends on the device models used to describe the device behavior. As semiconductor devices shrink, the need for accurate circuit simulation SPICE model becomes acute. The most important test structures in an IC manufacturing process are the devices themselves. It is thus imperative that these devices are accurately characterized so the most accurate model parameter set for the device under test can be extracted. The device models usually consist of a set of model equations that are either empirical or derived from device physics or a combination of both. Therefore, the design of integrated circuits is heavily dependent on circuit simulation, which needs compact device models. From the measured device characteristics SPICE parameters are extracted.

Parameter extraction is an integral part of compact modeling. The goal of parameter extraction is to determine the values of device model parameters that minimize the total differences between a set of measured characteristics and results obtained by evaluation of the device model. Several programs for parameter extraction are available on a commercial basis. Agilent offers an integrated circuit characterization and analysis program (ICCAP, 2004). ICCAP offers device engineers and designers a state-of-the-art modeling tool that fills numerous modeling needs, including instrument control, data acquisition, parameter extraction, graphical analysis, simulation, optimization, and statistical analysis. ICCAP runs on MS Windows 2000 or XP platforms and is easily accessible via Internet. Microsoft's Remote Desktop Connection, an application that recreates the desktop of a remote machine on the local machine over a network, may be used for parameter extraction from remote locations.

Silvaco's Universal Transistor Modeling Software (UTMOST, 2008) is another data acquisition and parameter extraction tool with applications in the areas of device characterization and modeling. UTMOST generates SPICE models for analog, mixed-signal and digital applications. UTMOST accepts data from direct measurements, measurements stored in a measurement log file, process and device simulation, or from other model parameter sets in order to create a suitable parameter set for a chosen model. UTMOST can control a wide range of commercial measurement equipment and probers so a user has maximum flexibility in the configuration of a measurement system. UTMOST supports the simulation of dc, transient, capacitance and s-parameter characteristics. UTMOST also incorporates commercial device models, user-defined device models and macro models.

THE TECHNOLOGY CAD COURSE

The TCAD course, offered since 1998 at IIT Kharagpur, has four primary objectives, and upon successful completion of the course, the students should be able to demonstrate a high level of proficiency. These objectives are as follows:

1. The use of TCAD tools allows students to learn the fundamentals of device processing in a virtual environment, including the determination of SPICE parameters;
2. The ability to develop skills on semiconductor process design (very similar to the actual experience obtained from a fabrication laboratory);
3. The ability to acquire proficiency with the use of TCAD tools and issues relating to analog and digital circuit design; and
4. For a career in circuit design (both analog and full custom digital) and/or as process engineer.

To support the aims and learning objectives more effectively, the course goes beyond the possibilities offered by traditional lecture type microelectronics courses. The curriculum was developed by researching various universities recognized for microelectronics education. TCAD software tools are introduced to students taking a required course in semiconductor device physics in the electrical engineering program. The course consists of theory (3-credits) in classroom (three hours) and the online TCAD laboratory (http://lod.iitkgp.ernet.in/netlab/). The classroom teaching is used to present and develop specific topics that support theory and the corresponding laboratory work. As part of the course, students are introduced to the TCAD tools during class demonstrations by the instructor. Students are expected to spend time in the classroom learning the theoretical background, objectives, and purposes of the simulation runs to be performed in the laboratory sessions.

Visualization is one of the biggest advantages of TCAD tools. During the process simulation, the evolution of actual cross-sections of the structure can be seen. TCAD visualization can also be brought to the classroom in different ways. Simulation results in a fixed graphics format can be integrated into the lectures, and the evolution of a simulation can be presented as a movie, e.g., embedded in a PowerPoint presentation. Tutorials are critical to the learning environment because they are designed with the purpose of enabling students to learn the material independently in an adaptive learning environment (learning at their own pace and areas of interest). Following the classroom teaching, the students continue with the tutorials, which are assigned as homework. TCAD tools in the laboratory are made available in an open laboratory environment in which students can use the software independently at any time and from anywhere. TCAD softwares (mostly the student versions) run with reasonable speed on any modern laptop under Linux or Windows operating systems, making interactive simulations possible in classroom as well.

The topics covered in lecture classes are the introduction and overview, role of TCAD for semiconductor technology development, TCAD principles, tool integration, process technology for Si, silicon-germanium (SiGe), III-V semiconductors, process simulation, device simulation, bipolar and MOS device models, heterojunction device modeling, simulation of SiGe HBTs, simulation of heterostructure FETs, simulation of AlGaAs/GaAs devices, introduction to virtual wafer fabrication (VWF) automation tools, device characterization and dc and ac SPICE model parameter extraction, TCAD calibration procedure, process compact models (PCM), and design for manufacturability (DFM).

TCAD LABORATORY

With the advancements in semiconductor manufacturing, it is becoming increasingly difficult for the VLSI designers to optimize the design without considering the effects of the VLSI processes. The conventional TCAD laboratory is designed for both undergraduate and graduate students. It takes place in a classroom where a stand-alone sever running the Silvaco suite manages twenty PCs as clients. The Silvaco software used by the students are ATHENA, the process simulator, and ATLAS, the device simulator. Prior to the beginning of the laboratory sessions, an interactive software presentation is conducted that helps the students set up the preliminaries necessary for simulation runs for the tutorials, like doping, diffusion, etc. The tutorial sessions also include several simple example runs for familiarization with the tool. The laboratory class consists of five/six sessions. All sessions are designed for developing two levels of knowledge: the definition/implementation of the simulation script (input file) and the analysis/understanding of the results.

Session 1: The process simulator ATHENA is introduced. The student simulate step-by-step an NMOS transistor. During this session, the doping profiles are investigated after each process step. In this session the effect of boron implantation used for the threshold voltage adjustments is studied. Also in this laboratory session the students observe the evolution of the device structure through all processing steps and are asked to write down their observations.

Session 2: Still with the ATHENA tool, this session continues with the polysilicon deposition and finishes the transistor processing. Like the prior session, the doping profiles are investigated after each step. The students are encouraged to compare the process flows for long channel and short channel devices.

Session 3: The device simulator ATLAS is introduced. In this session the students compute the ID-VD and ID-VG characteristics of the NMOS transistor for several process parameters. The gate sweep covers the operation regions of accumulation, depletion, inversion and deep inversion. The student then compares 1D cross-section of the carrier distributions at the center of the device with the theoretical predictions. With this comparison, the student is able to analyze the short channel effects observed in MOSFETs.

Session 4: In this session the mixed-mode device and circuit simulators are introduced. The electrical characteristics are used to obtain the compact model of the transistor and the students have to build a CMOS inverter.

Session 5: The student uses the Silvaco tools to design a process flow for a bipolar transistor using the graphical process flow editor tool. The students test the process flow by simulating it with the process simulation ATHENA and extracting the doping profiles across the emitter, base and collector regions.

Session 6: With the device simulator ATLAS, the students compute the output and Gummel characteristics of the BJT for several process parameters. The electrical characteristics are used to obtain the compact model of the transistor. The student extracts Gummel-Poon model parameters from TCAD device simulations.

TCAD Tools

During the last forty years, various simulation tools, mostly from the University laboratories have been reported. The principal aim has been to provide semiconductor engineers with computer simulation tools that form a quantitative link between the basic technological parameters and the electrical behavior of semiconductor devices. With the rapid development of the semiconductor industry, movement of individual TCAD tools from predominantly research and development to production environments has taken place (Fasching, Halama, & Selberherr, 1993). This fact has motivated a rapid development of TCAD systems as a uniform framework that supports the effective usage of multiple simulators with simple user interface, convenient and standardized data transfer, and visualization. Commercial TCAD vendors (Silvaco and Synopsys) provide an excellent collection of TCAD systems from different points of view regarding their respective origins, interests and goals. Commercial TCAD tools may be used to show students the link between physical and electrical simulation through the mixed TCAD and electrical simulation abilities of Silvaco and Synopsys tools. Advanced TCAD suites include process simulation, device simulation, compact models parameter extraction and circuit simulation suites, interconnect simulation, and optimization to other technology CAD requirements. However, the use and maintenance of coupled TCAD tools becomes difficult and requires a significant level of user experience.

TCAD Tools: Remote Access Solutions

Historically, all the TCAD tools developed were available on various UNIX-based platforms. Central computer servers often with multiple CPUs and UNIX- or Linux-based operating system and large amounts of RAM are used, as such servers are ideal for large-scale computing tasks such as TCAD. Attempts have also been made for Windows versions of TCAD tools, but the use of Windows versions are very limited, as the software packages are distributed and supported by third-party vendors. Also, as laptops are typically optimized for mobility and long battery life instead of large-scale computation, laptops are generally not used for running TCAD applications. However, the use of commonly available remote access software tools conveniently transforms a Windows-based laptop into a graphic terminal for a central UNIX- or Linux-based computer server. Choosing a remote access tool requires several important considerations, as they are available from various vendors and differ in several respects:

- Bandwidth requirements
- Support for encrypted data exchange
- Support for Open GL graphics
- Possibility to disconnect from and reconnect to a terminal session
- Possibility to use local Windows applications
- Ease-of-use and initial setup
- Cost

ONLINE LABORATORY AT IIT KHARAGPUR

Major steps for remote online laboratory development are (a) the design of the experiment, (b) remote control and operation of the instruments (for example, via LabVIEW, IC/CVlite, EasyExpert, VEE etc), (c) conversion to web applications, and (d) launching the experiments on the internet. A modular online remote laboratory typically consists of 12-15 hardware-based experiments that always need to be made available to students. The Laboratory-on-Demand (available at http://lod.iitkgp.ernet.in/netlab/) is an initiative to develop an online measurement-based novel integrated Microelectronics and VLSI Engineering Laboratory with simulation-based Technology CAD laboratory. The Home page for the laboratory is shown in Figure 1. An online TCAD laboratory gives the student, even at the undergraduate level, a chance to learn about realistic silicon wafer processing via hands-on simulation experiments. Recent advances in computer hardware now make even

Figure 1. The TCAD laboratory

a standard laptop suitable to run TCAD simulations in a matter of minutes. In Microelectronics and VLSI Engineering Laboratory, the students perform real measurements on a wide variety of devices. The measured experimental data are then passed on to the simulation-based TCAD laboratory for the extraction of SPICE parameters and will be discussed later in this chapter.

Microelectronics and VLSI Engineering Laboratory

Measurements are fundamental to an understanding of any semiconductor device. To teach device design and developing concepts at the Indian Institute of Technology, one of the premier engineering educational institutions in India, we use tools from the National Instruments (NI) electronics education platform. With the NI ELVIS integrated design platform combined with LabVIEW (Travis, 2000), we teach microelectronic devices concepts in the EC29004 Semiconductor Devices Laboratory and the EC39004 VLSI Engineering Laboratory classes. The EC29004 and EC39004 courses are required for juniors in the electronics and electrical communication engineering and electrical engineering degree programs at IIT Kharagpur.

The Internet is a powerful tool for engineers both for measurement and automation of the characterization laboratory and process optimization; allowing them to access from distant locations, along with their control. Internet-enabled measurement of devices utilizes the internet to achieve greater flexibility. The laboratory management system should be able to handle both the hardware and the software components. The hardware side includes the design and development of an experiment and controls it from the local computer. The software side covers the design and development of the tools for the physical measurement and administration. There are many factors associated with the network, the internet or the intranet, that have direct influences on the system performance throughout the remote operations. These factors among others need to be addressed for a proper application of online laboratories.

The Laboratory Server developed by the IIT Kharagpur NetLab developer team has several interesting features. Briefly, the IIT Kharagpur NetLab is an online Microelectronics and VLSI

Engineering Laboratory in which several experimental modules are included. In this chapter, we emphasize the device characterization module, which makes use of two distinct pieces of laboratory hardware/software. The heart of the system is the LCR meter and a semiconductor device parameter analyzer (Agilent E4980 and Agilent 4156C). Broadly speaking, the units generate the user-specified inputs to a connected test device, perform the test and collect the measured data. The Laboratory Server, at a high level, consists of three main software layers. The first is the Web Server Layer (WSL). This portion of the system contains all functionality that operates within the web server process space on the Laboratory Server. The second is a system Data Layer, which contains system information such as experiment logs, accounts, system settings, etc. Third, the Experiment Execution Layer interfaces directly with the laboratory hardware and is responsible for performing experiments submitted to the system.

Online Laboratory Development

During development of the Microelectronics and VLSI Engineering Laboratory, the roles of the System Administrator, Laboratory Manager, and students were considered in an integrated manner. For example, System Administrator's role includes (a) Design of online laboratory system specifications, (b) Design of measurement, monitoring and control interface, (c) Administrative interface and database, (d) Web server configuration design, (e) Design architecture for sharing remote laboratories, (f) Adding/removing experiment, (g) Failure detection/recovery, (h) User login, (i) Time scheduling, (j) Managing students accounts, (k) Automation of student performance evaluation, and (l) Student feedback. The Laboratory Manager's role includes (a) Preparing experiment modules, (b) Assisting students to conduct experiment, (c) Maintaining the experiment, (d) Developing new experiments, (e) Student's progress report generation, and (f) Laboratory grade sheet preparation/publishing. The students' role includes (a) Selecting laboratory and the experiment, (b) Studying laboratory manuals, (c) Appearing for preliminary quiz, (d) Booking experiment time, (e) Performing the experiment, (f) Passing the control to partner (collaborative learning), (g) Submitting laboratory report, (h) Appearing for viva-voce examination, and (i) Closing the laboratory session.

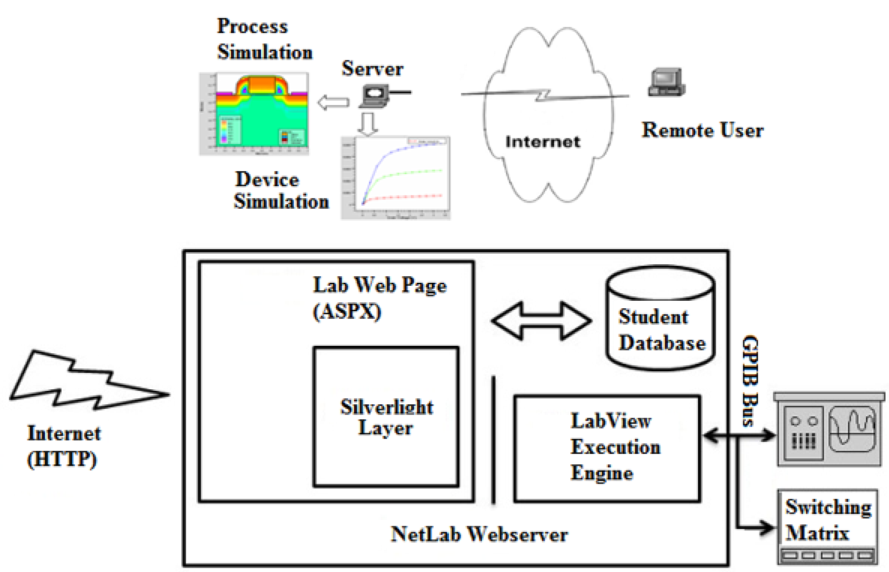
The System Administration Interface (SAI) contains Microsoft Silverlight or equivalent Web Application that captures the administrative functionality necessary to run the Laboratory Server. This includes human interfaces for log/record viewing and resource management. This interface is accessible only to registered Laboratory Server administrators and, as such, also contains authentication and authorization functionality. In the current version of the Microelectronics and VLSI Engineering Laboratory software, the validation of experiment specifications is handled entirely by the Laboratory Server. The experiment validation component contains all the functionality required for implementing any experimental module. After the experiment module is confirmed, the specification is checked for permission for submitting the experiment as well as the software imposed (by LabVIEW) constraints on the device. This confirms that the specified experiment should be allowed to run within the current configuration/session.

The Experiment Execution Layer (EEL) is a laboratory module (experiment) specific set of components that interfaces directly with the laboratory hardware and governs the experiment execution process. This layer can be broken into three individual components: the execution engine runs as a separate and independent process and interacts with the other Laboratory Server layers. The execution engine runs in the background at all times while the Laboratory Server is operational. At a given interval, the execution engine checks the experiment execution queue for new jobs. In order to communicate with the laboratory hardware, the Experiment Execution Engine (EEE) uses the methods defined by the set of Hardware-Specific Drivers (e.g., Agilent E4980). This component is made up of a series of custom built drivers specific to the hardware. In particular, each of these driver libraries defines methods that are specific to the piece of hardware they are meant to control using low level GPIB commands. At the lowest level of the Experiment Execution Layer is the library defining the GPIB Interface API. For the Laboratory Server to control the necessary laboratory hardware via a GPIB network, a GPIB Interface card supplied by the manufacturer must be installed. Combination of National Instruments (NI) software and hardware is found to be very effective for development of web-based laboratory such as Microelectronics and VLSI Engineering Laboratory. With the flexibility and power of NI ELVIS and LabVIEW, students can perform automated measurements. The online laboratory organization is shown in Figure 2.

Laboratory Management

As one particular experiment can be performed at a time by an individual student or a group of students (in case of collaborative learning), a careful scheduling of the experiments is required. Thus, for proper implementation and management of online laboratory system, an efficient laboratory management system is essential. As the experiments require 'real-time' control of the equipment, the online laboratory currently uses 'time-slot', during which the students set the measurement conditions and conduct experiments from remote PC. The time-slot (duration) is the time necessary to carry out an experiment and needs to be pre-estimated. The scheduling system for online experiments assumes the student would be able to finish the experiments during this time, just like they would in a conventional laboratory. The students first select a time slot and register (booking) to use the laboratory. At the scheduled time, the student logs in and is able to launch the laboratory client. However, the students can read manuals, watch experimental setups and videos, analyze the measured data without occupying the measuring equipment. Interactive experiments require the control of instruments, during which the users set the parameters and observe the results. As the actual measurement time is usually very

Figure 2. Online laboratory architecture



short, it is not desirable to provide the equipment to students for long periods of time. In fact, for a measurement, most of the time is spent setting up measurement conditions and post experiment data analysis. Thus, for proper utilization of the resources, proper scheduling is necessary in the laboratory management system. In order to choose the laboratory service provider with different kinds of measurement equipment and target devices, an efficient equipment sharing software tool is necessary in case of multiple student requests arriving at the same time.

When a student is prepared to conduct the experiment involving a unique experimental setup (i.e., cannot be used concurrently by others), the student logs in and is able to launch the laboratory client for the time slot already registered. Each experiment is associated with a given time length the administrator thinks is enough for the experiment, keeping in view the speed of the Internet and other network elements. The workflow may be summarized as follows:

- Selection of the experiment
- Opening of the scheduling web page link in the experiment page
- The system displays the scheduling interface for the next 15 days
- The system displays the available time slots for the selected experiment in green color
- The learner selects the date/time for an experiment in the scheduling interface
- The learner selects the desired time slot and submits the request for booking
- The system saves the learner schedule information and updates the scheduling database.

The system displays a confirmation of the schedule to the learner. In due course, the user is given a temporary link to the web page through which the equipment can be accessed and controlled. After the specified amount of time, a JavaScript automatically closes the window with a warning. Some of the advantages of the time-slot scheduling are the user gets full control of the equipment/system and performs the experiment at any time he/she wants and does not have to wait, and it is ideally suited for experiments where several measurements are necessary under different measuring conditions. However, the utilization of the resources is poor and the efficiency of the management system is not high since the equipment is not always in use and is sitting idle for some time and the number of users accessing the resources is low. Existing remote laboratories limit the number of users to the number of actual available resources. By integrating mixed-mode scheduling services in the laboratory management system, it is possible that a diverse set of experiments can easily be deployed online, and resources can be shared, thus optimizing the resource utilization.

MOSFET Characterization and Modeling

In this experiment, the student first measures the MOSFET characteristics, such as output and threshold voltage characteristics. Using these measured data, the student extract the "MOS Level 1" compact model parameters for a long-channel device. For a short channel device, the students use TCAD simulations for parameter extraction.

MOS Capacitor Characterization and Modeling

Maintaining the quality and reliability of gate oxides is one of the most critical and challenging tasks in any semiconductor fabrication. Electrical characterization and monitoring are critical for maintaining gate oxide uniformity. Many electrical characterization techniques have been developed over the years to characterize gate dielectric quality. However, the most commonly used tool for studying gate oxide quality in detail is the Capacitance-Voltage (C-V) technique. C-V test results offer a wealth of device and process information, including bulk and interface charges and many MOS device parameters. The importance of C-V measurement techniques is that a large number of device parameters can be extracted from the high frequency C-V curve. These parameters can provide critical device and process information. In this experiment, 1D MOS capacitor model is used to explain the formation of the inversion layer in an MOSFET device. While such a 1D model is quite appropriate to describe the long channel devices, modern deep sub-micron MOSFETs exhibit strong 2D effects, such as charge sharing between the source and drain depletion regions and the depletion under the gate. With TCAD simulation, these effects can be visualized.

For the TCAD simulation laboratory, a physically based, semi-analytical, charge control model for Silicon-Germanium (SiGe) quantum well MOS capacitor structure has been developed for the purposes of simulation. The program calculates the capacitance, charge distribution in the SiGe quantum well and parasitic surface inversion layer under various gate biases. The SiGe layer should be underneath a thin Si cap layer to assure formation of a high quality gate oxide since oxides formed directly on Silicon-Germanium have high density of interface states. The MOS capacitor structure consists of the following layer sequence: (100) oriented p-type substrate, p^+ doping spike, Si spacer layer, SiGe quantum well, Si cap layer, oxide and gate. It is assumed that the semiconductor region is doped uniformly throughout with N_d donors/cm^3, with the exception of the p^+-doping spike, which is doped with N_a acceptors/cm^3, and thin enough so that it is fully depleted at zero gate bias. Here, we have utilized quantum mechanical descriptions for mobile charge distributions in the Silicon cap layer and SiGe well layer using internal electrostatic potential calculation. In this experiment, the students first measure the MOS capacitor characteristics and then compare the experimental data with TCAD simulation.

Bipolar Transistor Characterization and Modeling

In the following, we discuss in detail the bipolar device characterization module, which makes use of a single instrument (Agilent 4156C) for the characterization of bipolar junction transistors (BJT) involving five experiments sequentially. The Experiment Module for BJT characterization consists of the following experiments:

- Static Collector Characteristics
- Gummel Plot
- Emitter Resistance
- Collector Resistance
- Current gain vs. Collector Current

In this experiment the students are asked to do the following:

- Collect data on output and transfer characteristics for a given device,
- Graph observed characteristics
- Download and analyze data
- Extract device parameters from the data and model the bipolar device

The complete range of semiconductor dc parameters can be quickly and accurately evaluated with 4156C stand-alone instrument. The measured static collector characteristics of an NPN transistor are shown in Figure 3. Results may be used to analyze to obtain early voltage and collector output resistance. One of the most important steps in evaluating bipolar transistor parameters is measuring collector current and base current as a function of base-emitter voltage (Gummel plot). These measurements can be graphically analyzed to obtain saturation current, current gain, and current gain vs. collector current characteristics, along with base resistance and recombination current characteristics. The collector current and base current characteristics are illustrated in Figure 4. The series resistance of the emitter of a bipolar transistor can be determined by stimulating the base with current and measuring the voltage between the collector and emitter. The inverse of the characteristics curve gradient can thus be obtained for emitter resistance determination. The measured characteristics are shown in Figure 5. Figure 6 shows how the 4156C can be used to accurately measure series collector resistances in low voltage regions, making it particularly valuable in parametric analysis of CAD models. The measured transistor's current gain vs. collector current characteristics is shown in Figure 7.

In the integrated online TCAD laboratory described above, simulation program, Tool for Electronic Model Automation (TEMA) is used for the extraction of SPICE parameters (TEMA, 2010). The software is an EDA tool for automated SPICE modeling of advanced semiconductor devices for development of an Analog/RF circuit design kit. The tool is suitable for modeling of dc, ac, RF, 1/f Noise and high-frequency noise for on wafer or packaged semiconductor devices. The parameter extraction for the industry standard compact models, such as, SGP, VBIC, HiCUM, BSIM, PSP and HiSIM can be performed for design of complex analog/RF circuits and to verify analog mixed-signal designs. In the present version used in our laboratory, only the dc simulation is implemented. The extracted SPICE parameters from the measured Gummel plot (see Figure 4), are shown in Figure 8.

STUDENTS' PERFORMANCE EVALUATION

Keeping track of each learner's activities and progress in the laboratory sessions is essential for the instructor. In the Microelectronics and VLSI Engineering Laboratory, the students' performance evaluation (for each experiment) is generally done at several stages: i) pre-laboratory quiz (after which a student is allowed to book time and perform experiment), ii) submission of the laboratory report after the successful completion of the experiment, and iii) appearing in the viva-voce examination for the experiment. Once an experiment is successfully performed, the student needs to save the data for analyses. After the completion

Figure 3. Output characteristics

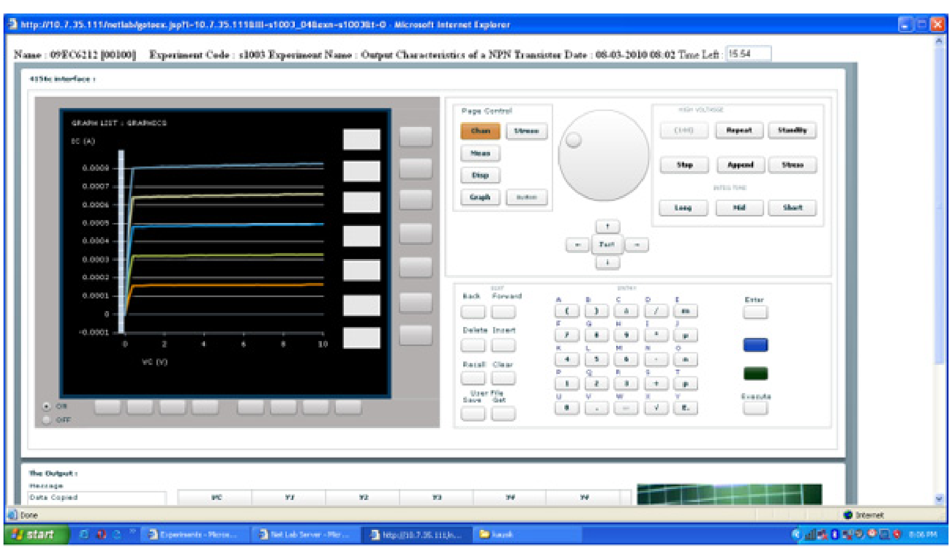
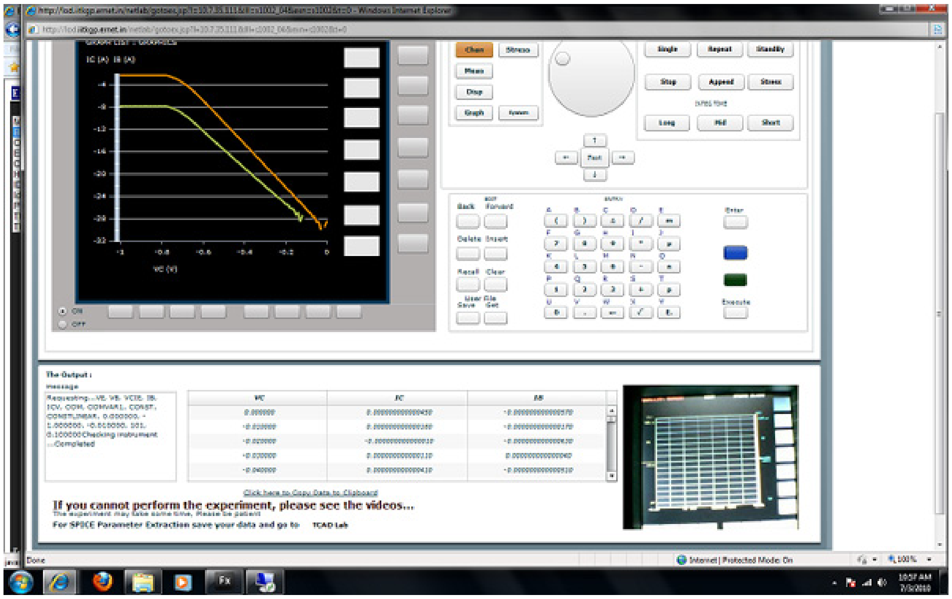


Figure 4. Gummel plot



of the data analyses, the student has to prepare a detailed laboratory report for submission. In the developed laboratory management system, an online submission facility is provided. The online report submission format contains the details from the experiment as well as the data obtained and the analyses. The uploaded laboratory report may be edited by the student for some time even after submission. Only the laboratory teacher has access to this report for evaluation and comments. Students are not allowed to have access to reports submitted by other students. Only after successful submission of the laboratory report, can the student take part in the viva-voce examination,

Figure 5. Emitter resistance

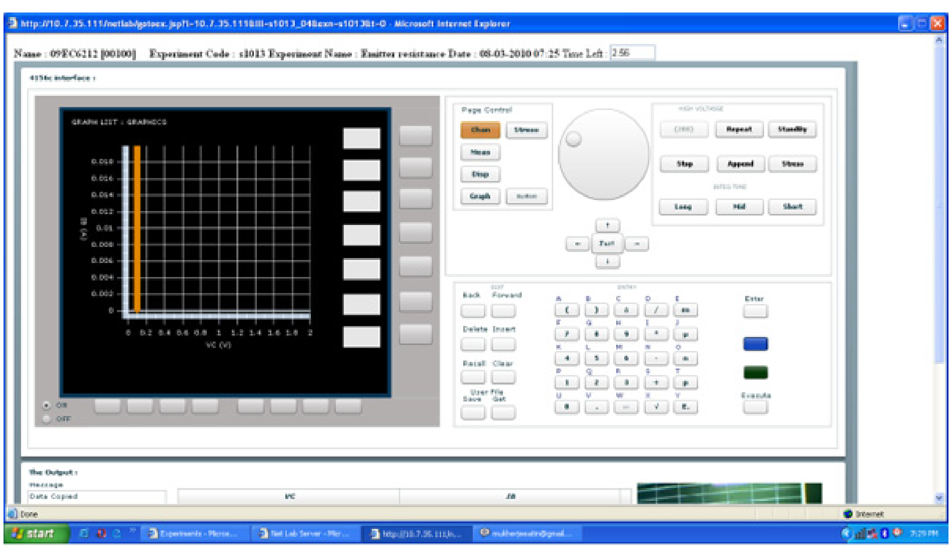
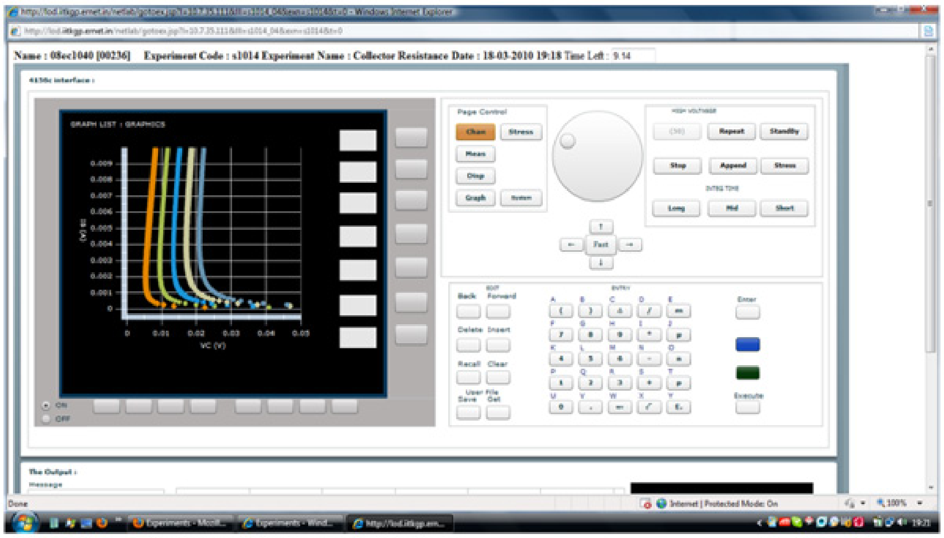


Figure 6. Collector resistance



which consists of several online true/false, yes/no or short questions. Both the viva-voce performance and the laboratory report are checked for each experiment performed successfully by a student. Once the student completes the laboratory module (say consisting of 10 or 15 experiments), then the final grade is prepared and displayed. Only the student and teacher can see this grade. Finally, for all the laboratory users, the final grade sheet is prepared and displayed.

Preliminary Quiz

The main motivation of online hardware-based laboratory experiment is to provide the facility to as many students as possible, and thus the number

Figure 7. Current gain vs. collector current

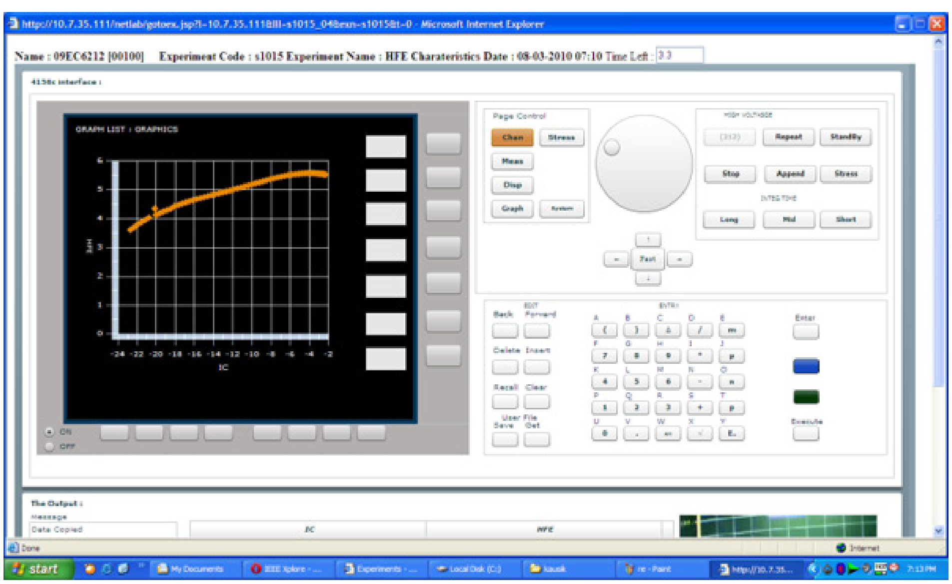
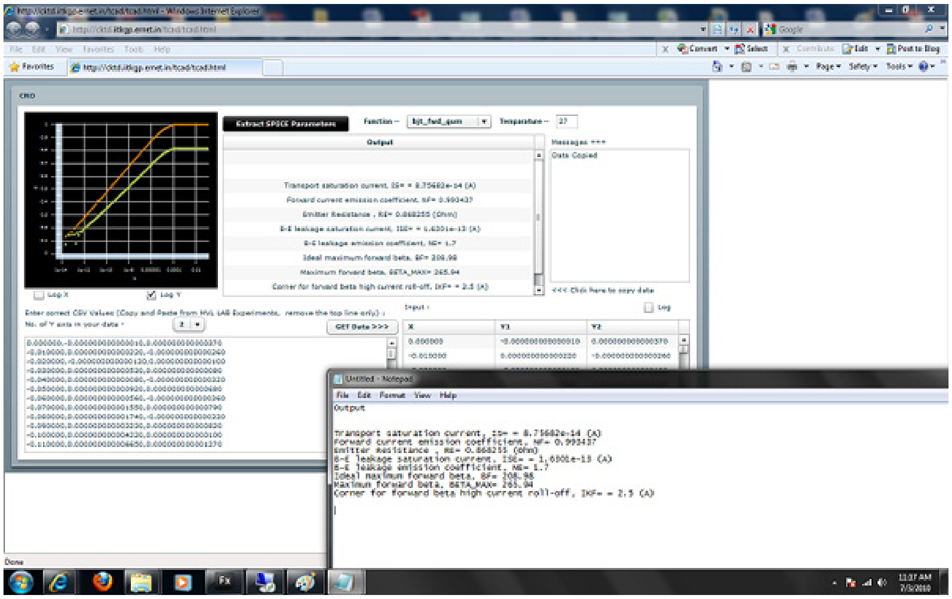


Figure 8. SPICE parameter extraction



of users per instrument is large. It is thus necessary that a user really engages the equipment only when he performs the experiment. This is why a preliminary quiz is introduced and conducted before the learner can gain access to the actual experiment webpage and the equipment and/or the experimental setup. This is done primarily to optimize the use of the instruments by making sure that the user is familiar with the instruments and has sufficient knowledge about the experiment. The quiz consists of several random true/false questions about the instrument, setup, experiment and fundamentals, etc. The user goes through the learning materials, videos, and details of setup available for each experiment. When the user is ready to use the instrument, he/she appears for the quiz. Only after the learner obtains a certain mark (set by the instructor for each experiment), is the student allowed to book the experiment time slot and gain access to the experimental setup during that time.

Uploading Laboratory Report

After the experiment has been successfully performed by the learner, is the experimental data saved and analyzed by the student for the preparation of the laboratory report. This report should contain detail from the experiment, setup, data, and results for parameter values, etc. The online laboratory report submission is basically filling in a form. After the preliminary section of the laboratory report is submitted, the user needs to enter the final results (parameter values) obtained from the analyses. These parameter values are then compared with the pre-set (default) parameter values generated by the instructor. Depending on the off-set (in percentage), the student is given marks, and grading is done accordingly. The instructor needs to set some tolerance for evaluation purposes and the system compares the student's uploaded value with the default value. For example, for a tolerance of 10%, 1 mark will be deducted from the full marks (FM) for the parameter. The instructor may also specify 3 tolerances levels and the corresponding marks that are to be deducted.

Viva-Voce Examination

Finally, the learner has to take part in the viva-voce. The viva-voce consists of several true/false or multiple choice questions as asked in a conventional oral examination. The students need to answer them in a certain period of time. The data is matched for correctness, and the number of correct answers is determined. The performance is graded accordingly. This value is then added to the laboratory report marks to get the final marks for the experiment.

Final Grading

The final grade for the student is prepared when all the experiments in the laboratory class have been successfully completed. Both the student and teacher can see the grade. The instructors and/or teachers can save this information for future reference. Marks for each experiment are then added to get the final grade report/sheet of the entire group of students for the class.

LABORATORY EVALUATION AND STUDENTS FEEDBACK

In order to learn how the remote laboratories provide useful learning environments, different models of assessment are employed. Our experience with internet learning systems has shown that the greatest benefits come from automating the laboratory assessment, as described in the previous section. To evaluate the utility of the online TCAD laboratory, we offered the students (with 27 students) an option to repeat the experiment in a scheduled laboratory class to improve their learning. Only 16 students operated the equipment for 20 minutes or more in the scheduled laboratory class, while nearly all the students managed to complete the assigned task using the remote laboratory. Our trial with the remote access laboratory showed that students could perform experiments in a lesser time. Using the remote laboratory system, 80% of the students managed to complete the required tasks, remotely operating the equipment, collecting data and running the simulations. These results demonstrate that the remote laboratory has significantly extended the laboratory experience for most of the students in the class.

TCAD tools are critical components of microelectronics engineering education. Over years of extensive development and incorporation of simulations tools in our course have taken place. Our impression is that available TCAD tools have been used extensively by the students and the students have received first-hand experience of using the tools. In the course evaluation and assessment process, the instructors have validated the outputs following the given specifications to assess the technical aspects of the home assignments submitted by the students. The students were asked to answer an online questionnaire and identify themselves by their roll number so that the instructor could relate their responses to logfile records of experiments performed. In a survey among the students on the use of the integrated Microelectronics and VLSI Engineering Laboratory along with simulation-based Technology CAD laboratory tools, several generic and concept-specific questionnaires were asked. Analyses of some of the responses indicate a good balance of available tools in the minds of the students. The assessment results collected from the laboratory class (106 students in two semesters) are analyzed in Table 1.

CONCLUSION

The Virtual Wafer Fabrication (VWF) has become an integral part of the semiconductor industry. The possibility of teaching semiconductor manufacturing in a university environment in a highly cost-effective manner by taking the advantages of high speed internet and available TCAD tools has

Table 1. Statistics of laboratory assessment

Evaluation Questions	Very Useful (%)	Useful (%)	OK (%)	Neutral (%)	Not Useful (%)
Adaptability	67	21	11	01	-
Effectiveness	73	12	06	-	09
Understanding of Experiments	57	23	11	04	05
Improvement in average student scores	23	49	21	-	07
Impact of the tool in terms of learning	59 (high)	31 (medium)	09	-	01
Scalability of the tool	66 (high)	24	10	-	-
GUI (user friendliness)	79 (high)	14	06	-	01
Online lab intuitiveness	43	37	11	09	-
Simplicity and interactivity	89 (high)	-	-	11 (medium)	-

been explored. Technology CAD is shown to be an excellent resource for teaching microelectronics. The objective of a laboratory component of any semiconductor fabrication course is to teach the students the unit processes involved in microelectronic fabrication and to introduce the practice of process development. Technology computer-aided design (CAD) tools for process and device simulation are now widely used in semiconductor process design. Microelectronic courses can be taught efficiently and cost-effectively by using process and device simulation tools without any physical process laboratory setup. When used in conjunction with conventional modes for teaching design and fabrication of semiconductor devices and circuits, the process simulation tools, combined with the device simulation, can be a very powerful tool for greater understanding of electronic devices and their operations. Device modeling and parameter extraction can help students understand the principles of electronic device operations and their applications in electronic circuits. With online the TCAD laboratory, the students are able to perform simulation experiments and explore the impact of process flow modifications at virtually no cost. An online Microelectronics and VLSI Engineering Laboratory integrated with a simulation-based Technology CAD laboratory can offer a new degree of flexibility in microelectronics education. However, the best environment for teaching semiconductor processing courses would be to use both the TCAD simulation tools along with the microelectronics laboratory, which is costly.

The developed online TCAD course with the associated laboratory gives the students, even at the undergraduate level, a chance to learn about real silicon processing via hands-on process/device simulation without the need for IC processing facilities and the opportunity to explore device behavior at a much deeper level. We expect online TCAD laboratories will encourage creativity, as interested student can design, create, and test devices in a matter of minutes to hours without using expensive resources.

REFERENCES

Dutton, R., & Yu, Z. (1993). *Technology CAD - Computer simulation of IC processes and devices*. Boston, MA: Kluwer Academic Publishers.

Fasching, E., Halama, S., & Selberherr, S. (Eds.). (1993). *Technology CAD systems*. Vienna, Austria: Springer-Verlag.

Fjeldly, T. A., & Shur, M. S. (2003). *LAB on the WEB, Running real electronics experiments via the Internet*. New York, NY: John Wiley & Sons. doi:10.1002/0471727709

ICCAP. (2004). *Agilent technologies, user manual*.

Kenrow, J. A. (2004). Integrating professional TCAD simulation tools in undergraduate semiconductor device courses. *Proceedings of the 2004 ASEE Conference*.

Maiti, A. (2010). NETLab: An online laboratory management system. *iJOE, 6,* 31-36.

Maiti, C. K., Chattopadhyay, S., & Bera, L. K. (2007). *Strained-Si heterostructure field-effect devices*. London, UK: Taylor and Francis.

Purdue College of Engineering. (n.d.). *Simulation Hub*. Retrieved from http://www.ecn.purdue.edu/labs/punch.

TEMA. (2010). Meridian software technology (P) Ltd. User manual.

The International Technology Roadmap for Semiconductors (ITRS). (2007). *The TCAD laboratory*. Retrieved from http://lod.iitkgp.ernet.in/netlab/.

Travis, J. (2000). *Internet applications in LabVIEW*. Upper Saddle River, NJ: Prentice-Hall.

UTMOST. (2008). *Silvaco International, user manual*.

ADDITIONAL READING

Maiti, C. K., & Armstrong, G. A. (2008). *TCAD for Si, SiGe and GaAs Integrated Circuits*. London: IET Press.

KEY TERMS AND DEFINITIONS

Architecture: The architecture defines a framework or structure of an online laboratory.

Client Requirements: The Client requirements describe different conditions (Operating System, Browser, Run-Time engines etc) necessary on a server/machine prior to run an application.

Device Simulation: Semiconductor device simulation is the modeling of semiconductor devices.

Experiment: Experiments are part of different types of Online Laboratories. Experiments may be real measurement-based or virtual (simulation).

Hybrid Laboratory: The Hybrid Laboratory is a mix of remote software simulations and real experiments in one single environment that can be accessed over the Internet.

Laboratory Management System: This represents a software application for the administration, documentation, scheduling experiments, tracking students' activities and performance.

Measurement Experiment: The structure of a measurement-based experiment is generally fixed. Measurement conditions and parameters might be changeable.

Microelectronics Laboratory: Microelectronics is a sub-field of electronics. Microelectronics laboratory is related to the study and manufacture of micrometer-scale electronic components usually made from semiconductors.

Online Laboratory: An Online Laboratory is defined as the hardware and software platform that provides resources over the Internet. Online laboratories may be divided in sub-categories:

Remote Laboratory, Virtual Laboratory and Hybrid Laboratory.

Process Simulation: Semiconductor process simulation is the modeling of fabrication processes of semiconductor devices.

Remote Laboratory: The Remote Laboratory provides real experiments that can be accessed via the Internet. This definition implies the control of real hardware and the realization of real measurements.

SPICE Parameters: SPICE parameters are semiconductor device parameters necessary for circuit simulation.

Technology CAD: Technology CAD is a branch of electronic design automation that models the semiconductor processes and semiconductor devices.

Virtual Laboratory: The Virtual Laboratory provides remote software simulations. The simulation model runs on the server and can be accessed over the Internet.

This work was previously published in Internet Accessible Remote Laboratories: Scalable E-Learning Tools for Engineering and Science Disciplines, edited by Abul K.M. Azad, Michael E. Auer and V. Judson Harward, pp. 185-205, copyright 2012 by Engineering Science Reference (an imprint of IGI Global).

Chapter 58
Implementing Business Intelligence in the Dynamic Beverages Sales and Distribution Environment

Sami Akabawi
American University in Cairo, Egypt

Heba Hodeeb
American University in Cairo, Egypt

EXECUTIVE SUMMARY

To compete successfully in today's retail business arena, senior management are often demanding fast and responsive Information Systems that enable the company not only to manage its operations but to provide on-the-fly performance measurement through a variety of tools. Use of (ERP) systems have been slow in responding to these needs, despite the wealth of the internally generated business databases and reports as a consequence of functional integration. The specific nature and demands by those senior management staff require the congregation of many external data elements and use data mining techniques to provide fast discovery of performance slippages or changes in the business environment. Data Warehousing and Business Intelligence (BI) applications, evolved during the past few decades, have been implemented to respond to these needs. In this case write-up, we present how the ERP system was utilized as the backbone for use by BI tools and systems to provide Sales and Marketing units in a transnational company subsidiary in Egypt to actively respond to the demands for agile information services. The Egypt subsidiary is the HQ of the African region's operations of several franchises and distributers of the company products, in addition to operating a beverage concentrate manufacturing plant in Egypt, which services the entire region's beverage products needs.

DOI: 10.4018/978-1-4666-1945-6.ch058

BACKGROUND

Company Overview

The case firm considered in this chapter is a transnational company subsidiary in Egypt, which is located in the Free Zone in Nasr City district, in the country's capital, Cairo. The company owns and operates a beverage concentrate plant within the domain of its facility for producing the concentrate syrup used in all its beverage products. These concentrates are then delivered to many bottling operators in African countries for the production of the final product mixing, bottling, packaging and trade fleet distribution in their respective territories. The concentrate plant is among the few plants of the transnational company world-wide, where the technical know-how formula of the company is produced, and the plant caters for the supply of concentrate syrup for the entire African, Middle East and Asian bottling territories.

The company is divided into a number of business units (BU). Egypt's subsidiary assumes the function of the head office to the African business unit which constitutes more than 25 franchises and distributers in many African countries. In this case-paper, we consider the analysis and evaluation of the information systems categories used in the head office of the company's subsidiary in Egypt for the Sales and Distribution management in the region. In particular, we detail how the backbone Enterprise Resources Planning (ERP) and business intelligence systems (BI) are integrated and used to leverage the need for agile management of the operations in the Sales and Distribution functions within the highly dynamic competitive beverage market.

Brief Economic Outlook of the Country Where the Case Company Operates

Forecasts put Egypt's food and drink exports growing by 59.4 percent between 2007 and 2012, which is not only a reflection of the free trade agreements ratified by the government of Egypt since late 90's, but also the country's improving food and drink processing industry. Regional trade agreements such as the Greater Arab Free Trade Area (GAFTA) have also given producers access to a far larger market. Having gone into effect in 2005, GAFTA has gradually lowered customs on locally produced food across a broad range of Middle Eastern countries. These agreements opened larger markets for Egyptian producers, given the similarity of diets and lack of language barrier. Meanwhile, Africa is also becoming another key export market, mainly for its proximity and lack of domestic production capacities in the African countries.

Sharply rising food prices have been the cause of growing unrest in Egypt over the past two years.

Though the Egyptian government has taken a number of measures to deal with the mounting public discontent, inflation hit the 20 percent mark and food prices skyrocketed, and a dire situation has evolved in the country.

Egypt is a rather unique market for the region as it benefits from a very large population (over 80mn) and an unsaturated food and drink market. However, the food and drink trade balance is highly dependent on imports. In addition, inflation, in a country with national poverty rate of 22.9 percent according to the World Bank Development Data Group, (2002) and Earth Trend Country profile estimating 20 percent of the population below the poverty line, led to increased levels of political risk and unrest. Against a backdrop of worsening global financial market turmoil and rapidly accelerating inflation, particularly in emerging markets, Egypt has significantly more economic hardship clouds on the horizon. Nonetheless, Egypt receives a pretty high score in the region for its food and drink market due to its per capita food consumption growth. Egypt does not fare well for the country structure indicator, with low GDP per capita, although the size of its population and lack of market maturity help pull up its score.

SETTING THE STAGE

Companies in the beverage industry often operate as producers and distributors of canned and bottled soft drinks, concentrate and other juices and liquid beverage products. They experience rapid changes in the way they have to do business. For one thing, increased commoditization and diversification of beverage products has made it more difficult than ever for beverage manufacturers to differentiate their products and their brands against rivals and competing brands.

In addition, beverage consumers are demanding greater product variety, higher levels of service and more value for their money (Bingham, 1999). The increased global competition, escalating retailer demands and increasingly stringent government regulations have added to the pressure on those companies, making it harder for them to compete globally. With growing concerns with food safety issues, these companies are also finding it necessary to monitor and track every phase of operation – from raw ingredients through finished product, to packaging, storage and distribution – to ensure compliance with the higher safety standards.

Driven by these challenges, beverage producing and distribution companies are looking more and more to their manufacturing processes and supply chain in an effort to increase production efficiency, reduce operational costs and effect better overall management of the enterprise and its assets to sustain desired levels of profitability and growth.

Business Imperatives

Historically, there has been resistance to the introduction of technology in the beverage industry except perhaps to automate certain production processes (e.g. mixing, canning, bottling and packaging). In many large companies, use of information technology (IT) has been limited to off-the-shelf or custom-tailored systems implemented to assist employees with such discrete business processes as accounting, corporate finance, human resources and purchasing. However, in the past few decades, many firms have implemented Electronic Data Interchange (EDI) and host of functional separate systems, often in response to customer and vendors demands or management directives for adopting business-to-customer (B2C) and business-to-business (B2B) e-business functionalities. Despite all of the information flowing through these disparate systems across the enterprise, the information services to operational and top management more often than not lacked integration, timeliness and comprehensiveness.

The lack of timely financial information, integrated with order and production flow would hamper the ability of firms to know where they stand from day to day, both from financial and operations point of views. In most companies, by the time financial information becomes available in a report, it's historical and too late to change course to correct problems. In today's fast-paced global economy and rapidly changing marketplace, the lack of integrated, real-time information to provide business and market analysis, business intelligence and decision support tools negatively impact profitability, growth and ability to act quickly to external marketing forces. Moreover, efficient order entry, materials and inventory control and fast turnover dictate the need for agile and responsive information systems to minimize waste and delivering efficient customer service. In market-driven industries, once the product is made, time is not nearly so critical – but storage, transportation and distribution are.

On the beverage industry side, we see an astonishing level of diversification fragmenting this industry. Where once there were a finite number of carbonated and non-carbonated soft drinks occupying the competitive landscape, today there is an almost infinite explosion of choices – from pure juice and juice blends, to sodas of every kind, waters, and host of beverages. Sales in this area are market-driven, so producers tend to rely on special packaging, competitive pricing, event-driven promotions and other attention-getting techniques to differentiate themselves. For these firms, forecasting and timing of product-to-market are crucial. Market-driven companies thus need to be able to put a product out on the shelves, see how it sells, then turn to adjust production to match demand.

The preceding characteristics illustrate that each segment of the beverage industry has its own challenges to overcome whether it competes in local, national, regional or global markets. However, to retain customer loyalty and grow market share, all beverage manufacturers and distributors need greater flexibility to respond quickly – whether to changing customer expectations, government regulations, or fluctuations in the marketplace resulting from seasonal changes in drinking patterns. In short, beverage producer and distributer companies are finding they have to increase the speed or decrease the lag time between when their products are made and when they are out on the market. To coordinate their production capabilities, manufacturers also need to increase their ability to handle different materials and product lines simultaneously. If they are to improve procurement and inventory control, they must find a way to accurately track both raw materials and finished goods. And, to build their company and product brands among customers and consumers, they must be able to ensure consistency of quality across product lines.

On the business end, optimizing plant operations can be a tremendous source of savings. Companies attaining high visibility and close control of key production elements opt to simplify complexity and better manage cash flows. They need to be able to forecast more closely to customer demands – avoiding either overextending or under-producing – to improve their return on corporate assets. The use of enterprise resources planning (ERP) systems becomes the natural choice for fulfilling such needs (Davenport, 1998). ERP systems operate on enterprise-wide domain to ensure integration among all business functions and result in one large single view of the company's data resources. If those systems are well matched to the business and well implemented, many benefits can be accrued such as: decreased reaction times, better logistics flow, increased responsiveness to market and customer changes and improved supply chain management (SCM).

Information Systems Backbone Implementation in the Case Firm

By early 2000, the case company presented in this chapter planned to acquire an ERP system to replace its aging legacy systems. The company looked for a proven technology platform for its business applications with strong functionalities in areas such as financial applications, procurement, order entry and fulfillment, planning and scheduling, inventory management and optimization, product configuration, flexible product costing, manufacturing, EDI capabilities, and support of multiple plants and/or warehouses and complex distribution system. To that end, a blanket, multilayer deal with highly reputed ERP vendor was selected and configured to run several business application modules to cut costs and share data through multiple distributed systems located in multiple regional datacenters, with secured global access. These applications are built around ERP, customer relationship management (CRM) and SCM modules as well as set of integration middleware tools. The system aimed at helping improve the logistics of store deliveries, sales orders and the back office operations of the networked company's manufacturers, bottlers and distributors around the globe. This distributed approach targeted the improvement of market execution, better service to consumers, in addition to giving a more integrated system platform to serve the information needs at the store level and account level, and more effective management of the business on the street. The implemented distributed ERP systems provided the following application and service tiers:

1. The back office tier which included several modules. The African region implementation provided the following modules:
    a. Marketing Expenses Management
    b. Business Travel Management (BTM)
    c. Human Resources (HR)
    d. Financials and Sales
    e. Logistics
    f. Supply chain and Manufacturing
2. Data warehouse and decision support tier which included the following modules to serve the African business unit:
    a. Value chain modeling, the module manages everything that adds value to a product. Starting from raw material used to produce the product such as sugar, water, vitamins and caramel, to the time employees spend on various functions and the salaries of employees, to the trucks used to distribute the product and the gasoline needed for the trucks. It uses all those individual costs to know how much the product should sell.
    b. Forecasting module, the sales business intelligence module is responsible for everything related to sales. It manipulates three types of data: historical data, which include the actual weekly sales; the business plan data, which constitutes the forecasting for the sales of the year; finally, the rolling estimate data, which is a more accurate prediction of the sales based on the discrepancy between the actual sales and the business plan.
    c. Decision support System, this module helps sales function managers in making decisions based on sales by providing the following services:
        - Sales flash report issued on specific day of the week. Sales reports of the week are sent to employees on company-provided blackberries. The weekly sales updates are given versus business plan predictions or versus prior years. In addition, sales is always directly related and translated into market share. The company used push strategy to alert the employee that he/she has received an email.
        - Telecom reporting is used to control the escalating costs of telecommunications within the African business unit. Managers get periodic data from mobile service operators to determine how much their units pay for mobile services. This particular application used the EDI facility provided by the system through the middleware layer.
        - Margin minder report is used to track sales in every outlet, and provides small outlets with data on how to make more profit. For example if a small café is buying more of a certain Stock Keeping Unit (SKU) but the reports show that the café is not selling much of it, the system then recommends that the outlet buys less of this SKU and buys another one that seems to be selling good at the particular café. In addition, it does the 'Right Execution Daily' (RED) that manages the product display and the picture of success (pos), i.e., to make sure that company logo is put at the entrance of sites, that the tempera-

ture of the product is set at the prescribed temperature, that the company refrigerators given to sites have their products only.

d. Business Intelligence (BI) tools which perform activities such as collecting and transforming data from a variety of systems, consolidates and aggregate these data in readiness for reporting to assist in decision making. A visual reporting component was missing from this battery of tools for presentation of knowledge to assist with decision. At this stage of implementation, as such, this tier did not incorporate dashboard to provide visual summary of the operations key performance indicators (KPIs). This particular shortage is the subject of the additional development addressed in this chapter.

Additionally, there are some applications used in the African region Egypt HQ that the IT manager referred to as 'ko' which literally means 'knock-out', but what it actually meant is that it is site-specific, internally developed by the company, for the network of sites in the African business unit

A middleware tier was deployed by the company to achieve integration among these disparate enterprise applications through using several reporting and communications layer comprising:

- **Mail:** the mail system used by the company is 'lotus notes' platform. This platform provides a collaborative work environment as well as email service.
- Microsoft office: the end-user productivity tool for the client side hardware (PCs, laptops, PDAs, etc.)
- **Share point portal service:** used as the gateway/ portal through which authorized employees can gain access to the various service provided by the host of applications in site or globally.

The architecture of the technology service layers after the implementation of the ERP system in the datacenter accessed by Egypt subsidiary HQ is shown in Figure 1. Some of the modules were implemented in a "vanilla" standard mode while others were customized to adapt to the beverage industry-specific operations. The company gets, through a blanket contract, upgraded versions of the modules on a periodic basis (for the standard application modules only).

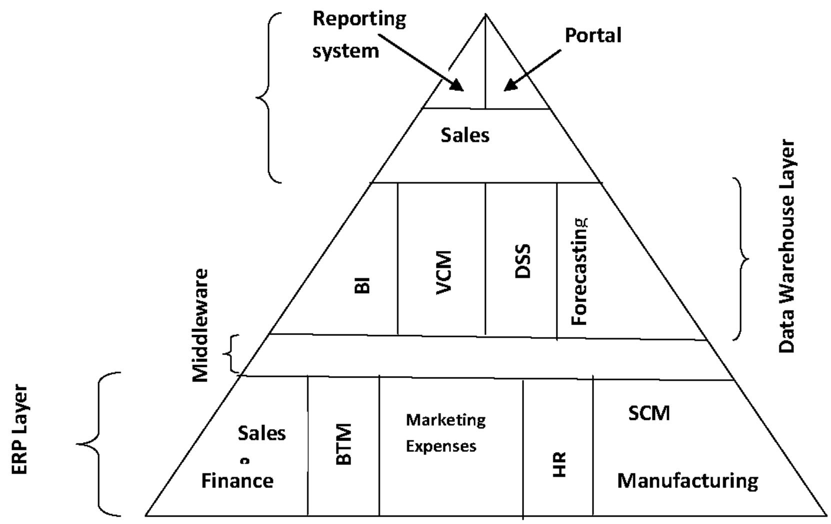
The network and communications infrastructure provided connectivity to the corporate datacenter for all sites in the African BU. Business transactions are entered on real-time basis into the respective ERP module. Each site must thus secure the health status of its internet/private connectivity and monitor its operational effectiveness and availability of its links.

The Sales Reporting System Prior Implementing the New Dashboard Facility

For each regional site, the specific reporting and operational needs (which included manufacturing supply chain, sales force management, marketing, inbound and outbound logistics), are accessible remotely from the regional datacenter of the company in Africa. Several functionalities were considered for the management of the African BU in Egypt's HQ site, which is concerned, among other business functions, with the consolidated sales within its domain. The process with which this objective was initially achieved is described in Figure 2. In this process, sales data exchanged between the franchises/bottlers and the HQ constituted:

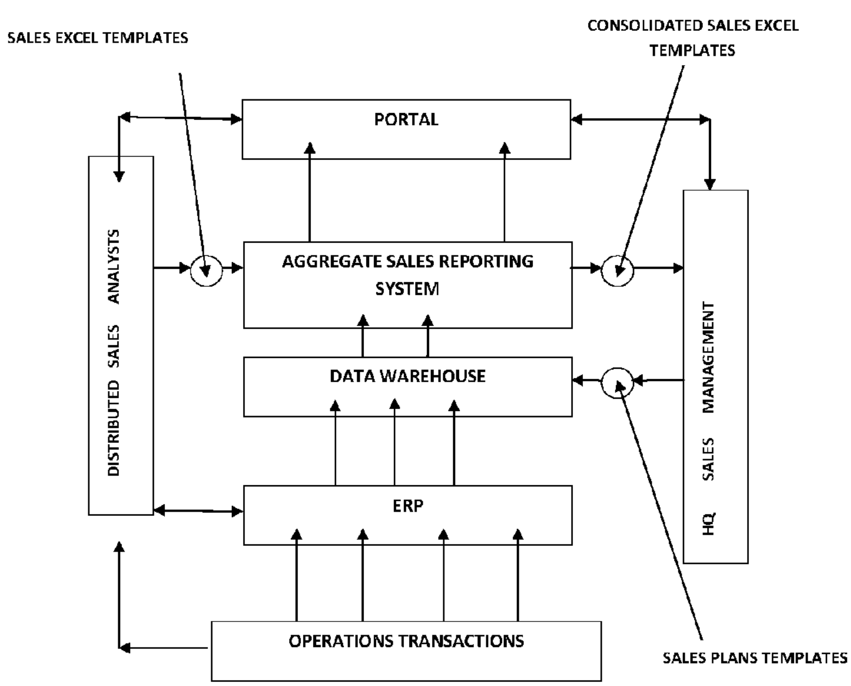
- **Actual sales figures:** These are the actual sales volume reported by the bottlers. Actual sales were communicated on a daily basis through e-mails, via Excel templates, sent to every franchise analyst.

Figure 1. Architecture of the IS layers used in the company before the new BI dashboard implementation



- **Rolling estimates (RE) figures:** which reflect the adjusted sales on a monthly basis to adjust targets during the course of a year, thus reflecting changes in the market and used to predict sales volume for the balance of year. These figures are confirmed after communication between the franchise managers, the business planning manager and the HQ office. Once confirmed, the business planning manager sends the final

Figure 2. Sales reporting process before implementation of the IDB system



confirmed version of the file which contains the actual sales so far and the balance of year figures on a monthly basis for all sites in the domain countries

- **Business Plan (BP) figures:** are set by the BU management at the beginning of the year. They do not change during the course of the year. The BP dictates target sales per brand per pack per month.

Sales Excel sheets Templates which are formatted by the sales analyst of each franchise/bottler liberally, are used by the company for sales reporting. These templates constitute the following reports:

1. Daily Reports

Bottlers of each country send daily sales reports that include the daily actual sales volume in both physical and unit case volumes to the franchise sales analyst. If they send it in physical cases, the franchise analyst converts it into unit cases using the unit case conversion factors. The reports include sales by brand and pack for every day of the week. The drawback of this reporting scheme is that there is no fixed template format for the daily reports sent by bottlers, who communicate the daily sales volume to the HQ.

2. Weekly Reports

The franchise Sales Analyst prepares and sends weekly reports after consolidating the daily sales sent by the bottlers to get total weekly figures. The reports also display the rolling estimates split weekly in addition to the weekly budget split and prior year weekly split. The Business Planning managers use the data communicated through those reports and consolidates data of all franchises to send consolidated weekly reports to HQ.

3. Monthly Reports

Every franchise analyst is requested to submit monthly reports as per deadlines dictated by a quarterly reporting calendar communicated by sales analyst. Those monthly reports include the following:

- Best estimate for current month, draft of upcoming month RE and upcoming month weekly split. This data is usually communicated before the current month is closed.
- RE volume by brand and pack. This report is submitted after the month is closed, it communicates the 'Actuals' of the closed month in addition to the RE figures of the balance of the year data. All 'Actuals' and RE figures are provided with a brand/pack split.

The HQ Sales Analyst then consolidates monthly brand/pack data sent from all franchisees to prepare brand/pack monthly RE updates.

As noted from this process, there are many countries involved in the reporting cycle of the sales results and many reports are communicated back and forth, using email and non-standardized Excel sheet templates between the country sales analysts and their corresponding managers in the HQ site. Many errors and inefficiencies continuously existed due to the use of those sales templates. The oft problems encountered included the following:

- Data inconsistencies due to inconsistency in product coding and Excel template variations led to lateness in reporting;
- Miscommunication due to network problems, variable semantics and other inconsistencies resulted in many duplicate data items (same data repeated in several reports);
- Consolidating non-uniform templates data and variable-semantics led to waste of time and efforts;

- Lack of analysis due to meager analytical competence of sales analysts;
- Lack of standardization led to too many data but no information;
- Redundant processes resulted in data chaos (too many templates causing confusion).

To that end, the focus of this case write up is to determine how the company utilized a new business intelligence dashboard tool, to provide the Sales and Marketing units with the necessary agile and fast decision making platform and to respond to the dynamic market and consumer changes over a wide geographic area; at the same time overcome the critical impediments illustrated in the process described earlier and schematically diagramed in Figure 2.

Literature Review

During the past two decades, many organizations adopted the ERP as their preferred backbone information system for data and information management, since then world ERP market poised to command a little under $65 billion by the end of 2009, according to AMR Research Market Analytix Report (2005). The use of ERP systems by those organizations targeting cost reduction, increase productivity, improve customer satisfaction and suppliers relations in the drive to improve competitiveness in the global networked market space (Davenport, 1998). The cost expended for the acquisition and implementation of ERPs exceed millions of dollars and affect the overall firms' earnings and revenues (Davenport, 1998) and their market values, (Chatterjee et al. 2002). With such huge capital investments, many firms are forced to assess carefully their return of investment (ROI) of such an infrastructure.

The impact of adopting ERP normally goes beyond the immediate control of the business resources, it sets the grounds for organizational changes and the way business processes are performed, (Kallinikos, 2004). As noted by Brazel and Dang (2008), the implementation and use of ERP systems represent a radical change from the operation of legacy systems. Many researchers have therefore studied ERP use and adoption from various perspectives. As a category of information system, several authors approached research on ERP from the perspective of information systems success model (DeLone and McLean, 1992, 2003; Gable, et al. 2001 and 2003; Seddon, 1997; Sedera, et al. 2004; Ifinedo, et al. 2006a and 2007; Robey, et al, 1999).

ERP implementation is a challenging endeavor for many organizations, spanning many functional areas (Yen, et al. 2002), demanding high level of coordination among all stakeholders and adjusting to changes and synergy in the procedural workflow by every business function. Use of such systems implies changes in the workflow of business processes (Kallinikos, 2004). To that end, success of the implementation projects and the critical factors influencing this success was studied by many researchers to investigate ERP implementation success factors, (Bancroft et al, 1996; O'Leary, 2002; Ptak, et al, 2000). Other studies aimed at identifying the critical factors impacting implementing ERP projects (Fryling, 2005). Factors such as those influencing successful implementation, from the operational (rather than technical) point of view were addressed by many researchers, for example, the business operations coverage of the package and the number of licensed users (Francalanci, 2001; Kumar, et al. 2001; Markus, et al. 1988; Parr and Schanks, 2000).

The impact of system configuration and/or setup revisions and enhancements was also studied by Fryling (2005), Nicolaou, et al. (2004 and 2007), Light (2001), Mensching, et al. (2004) and Nah, et al. (2001); impacts of organizational and national cultures (Krumbholtz et al. 2000; Soh, et al. 2000); the configuration and setup of the package's parameters tables and their overall impact on the architecture and flexibility of the ERP packages for rapid adaptation (Fan et al. 2000; Spott, 2000). Some articles also addressed the impact of ERP process standardization and the restructuring of organization tasks or business processes reengineering (Kumar and Van Hillgersberg, 2000). The ability of the system to provide quality and trusted information for operational and executive management decision making has been identified by many researchers (Xu, 2003; Madapusi, et al. 2007).

ERP integrates all functional areas of the organization, acting as the backbone of the information management platform of the organization (Chou, 2005). With its integrated database, ERP systems integrate other functional business components such as Customer Relationship Management (CRM) and SCM systems data resources to support informational needs for decision making. Since ERP system is internally looking, the need to accommodate access to data across the organization boundary is required for decision making at the strategic and organization-wide levels. It is not that ERP systems don't have wealth of information, they do; the challenges lie in the ways of mining them. ERP cannot facilitate real-time decision support function for several operational reasons. Since information is the foundation of every critical business decisions, Decision Support Systems (DSS) are vital for any organization (Drucker, 1998). Report writers can access data from multiple ERP modules and consolidate them with other data elements for decision support. Business Intelligence (BI) is about getting the right information, to the right decision makers, at the right time (Alter, 1980). It is an enterprise-wide platform that supports multi-dimensionality reporting, analytics and decision modeling leading to fact-based decision making and enabling to get a "single version of the truth" (Rasmussen et al., 2001). The common pain points that BI is used to solve are typical examples of what most organizations experience:

- Data everywhere, information no where
- Different users have different needs
- Excel versus PDF
- Pull versus push
- On demand – on schedule
- Your format – my format
- Takes too long – wasted resources/efforts
- Security
- Technical "mumbo jumbo" … Why I just can't get it to you when you want it.

By integrating business intelligence (BI) tools and ERP modules, data flows directly from the ERP database on real-time basis. However, some reliability, availability and scale efficiency may arise as a consequence – particularly due to the excessive access load that may hinder transactional operations. Separating the active ERP database from that of the BI resulted in embracing a second data storage tier, a data warehouse (McDonald et al., 2002). The ERP-BI, On-line Analytics Processing (OLAP) and DSS tools integration framework is based on congregating all needed data from the ERP system and other external data resources, load them into a Data Warehouse or a data Mart, then link to several BI tools, such as OLAP, data mining, analytics tool and reporting systems to create more consistent and knowledge-centric data reporting. BI tools provide such functionalities. More and more organizations extend their ERP beyond the level of back-office to improve sales, customer satisfaction, and decision making (Stedman, 2002).

Integration of the BI and ERP system contributes additional values to businesses (Chou, 2005). According to Holsapple and Whinston (1996), characterization of common DSS features, BI generates different views for available data systems, a scaled data mart or data warehouse providing rich, timely and well structured and cleansed information to the BI. Bolt-on BI systems are also used to view financial, marketing and sales queries by using different tools (CRM2day.com, 2004). Customer Experience Management (CEM) in retailing has also been researched within the framework of BI by score of researchers such as Ding, et al. (2006) and Kamaladevi, B. (2010).

Issues of integration of SCM with customer experience management were dealt with by Chou (2005). Approaches to building and implementing BI systems were investigated by Celena Olszak, et al. (2007) being an important decision to many implementing organizations.

CASE DESCRIPTION

Technology Concerns and Needs

Within the context of the studied case, a new Sales Reporting process was developed to address the inefficiencies and pitfalls of the template based sales reporting system described earlier. Egypt's HQ IT department together with the Sales unit, embarked on the development of a new BI tool which was code-named the IDB (information dashboard). The new IDB tool was designed to provide enhanced method for uploading sales data, address the inconsistencies of the country sales figures and the cumbersome method of obtaining the consolidated reports at the HQ. Analysts were trained to upload data on the new IDB platform using a newly developed product coding system. They get the codes from commonly accessible codes database. The coding system is based on a code consisting of 10 digits. The first four digits represent the brand type and flavor. The last four digits represent the package size and type. The two digits in the middle represent the type of syrup whether liquid for normal packages or powder supplied to machines. This level of breakdown for any product provided an easy and efficient facility for tracking Stock Keeping Unit level since it embedded all the necessary information regarding the product whether in terms of brand type, flavor, package size, type, format……etc.

The implementation of the IDB tool availed single unified interface for users to upload the sales data from the respective country to provide aggregate reporting of sales for the HQ in Egypt. The tool allowed drill-down granularity for users. We will first analyze the structure of this implemented information dashboard tool (IDB) and then describe the new sales reporting process. The IDB acted as the single "database" view for all trade related information used by all franchisees and is used through applying several data mining techniques. It is a customized web based application built to access the warehouse database and allowed access from anywhere through the company's network. Furthermore, the IDB is used as the reporting vehicle to management - globally, retiring the use of once prevalent sales templates and the use of the email for posting these templates. Access to the IDB is achieved via web browser using specific URL or via the Company Portal (based on MS Share Point platform). Users log in through their IDs and passwords. After updating the system with the daily, weekly or quarterly sales, figures, respective users can log in and retrieve any report needed at any granularity level for any specific period.

The sales reporting process constituted several activities. Every franchisee analyst is to prepare the weekly sales data using the relevant codes and upload the data on the IDB. The upload is performed automatically through an Autoloader and scheduled on Mondays every week. The system decodes the uploaded codes and maintains the actual data on the system. Business plan figures and rolling estimates figures are also uploaded in the same format and manner. Whenever any discrepancy in the uploaded data on the IDB of a single country affect the accuracy of the data of the total business unit, control measures are promptly carried out to assure the integrity and accuracy of data on the system. After all franchisees upload their weekly figures on the IDB, the HQ Sales Analyst verifies the data on the system and adjusts/deletes any discrepancies found by detecting the countries responsible for the discrepancies and suggests methods for resolving these conflicts with the concerned analyst. The Sales Analyst can also report any variance detected on the IDB due to any error in the application engine through Business

Applications Call Logging system, used by the IT department as the support tool for any encountered technical problems while using the IDB.

The HQ sales analyst is also responsible for planning and coordinating all IDB activities with all franchisee analysts in terms of training analysts, installing new IDB applications and communicating with IT to resolve encountered problems. The HQ sales analyst is also responsible for issuing new codes for newly launched packs that are not on the system by filling a request form. The HQ Analyst also monitors the status of expired codes and de-activates them.

Developing the IDB Dashboard Layer

The described new sales business process is depicted in Figure 3. Upon upload of the sales data on IDB, users can simply log into the system from anywhere and access the reports menu via the portal or the specially provided URL and pick the type of report required using drop down lists. These lists include all product parameters. The IDB system enabled users to get any type of information by selecting the specific criteria needed in the downloaded report with any degree of granularity. In other words, users can customize the reports retrieved from the system by choosing the appropriate parameters from the displayed scroll down lists. Drill down lists may include, but not limited to, parameters such as:

- Country
- Bottler
- Period (Year/Month/Week)
- Brand
- Flavor
- Package type
- Pack size
- Historical data
- Business plan data
- Rolling Estimate (RE) data
- Analysis of sales performance versus prior years, BP, RE…..etc.

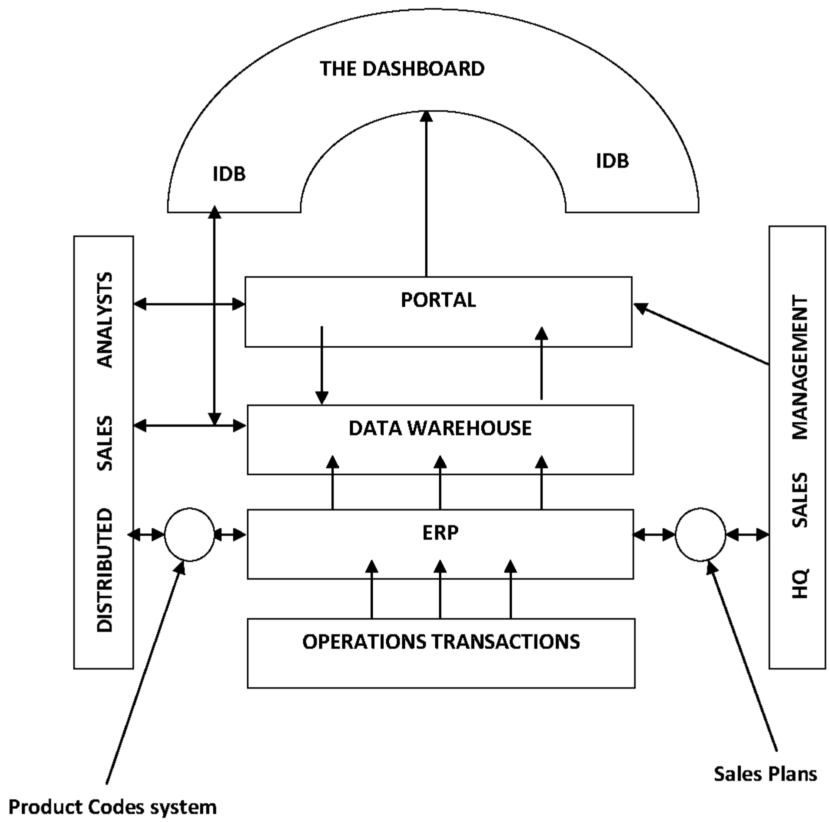
Access to the facilities offered in the IDB system is granted though a robust and premeditated authorization scheme. IT unit provides authorized users with their respective personal user name and password for authentication, however accessing and retrieving information at various data-granularity levels are subject to authorization policies. This authorization scheme is set by the respective country management, and communicated to the IT department, to provide various access privileges for the various organizational hierarchical levels.

The Impact of the IDB System on the Effectiveness of the Sales Reporting Process

The use of the IDB facility as a component of the battery of BI tools of the Company's IS platform enhanced the Sales reporting cycle in many ways and had increased the overall efficiency of the Sales Reporting Process. The major areas of improvement IDB offered are:

- First of all, it has eliminated the redundant process of sending the former sales reports templates that were subject to various interpretations, wasting time and effort and causing inconsistencies and mistrust of data.
- Using IDB as an outer shell of the data warehouse provided a very useful tool as it integrates all types of information required whether historical data, brand information, package information, business plan figures…etc. Before using IDB, managers and analysts had to go through many Excel files and sheets to obtain information which is time consuming and some of the required data were lost or hidden. Besides, the same information was sometimes obtained with different values from different sources causing confusion.

Figure 3. Sales reporting process after implementation of the IDB system



- Provided multi-dimensionality reporting using "slice and dice" operations by the analysts.
- IDB acted as a timely and trustworthy reporting tool with a single source of sales data which is not subject to various interpretations. All analysts are bound by specific deadlines to upload their data, so the managers are confident to find the data on time and to retrieve it in a standard format to facilitate the data integration and the decision making process.
- IDB provided analysis facility. Reports retrieved facilitated comparisons and analytical measures that can be used to help managers in the decision making process.
- IDB helped in determining profitable trading partners through the use of multi-dimensionality data mining
- Augmented and replaced cumbersome previous spreadsheet-based system.
- Enabled driver-based planning to streamline and focus efforts around HR plans, manufacturing requirements, or sales resources.
- Enabled use of rolling forecasts to increase forward visibility.
- Reduced the need for consolidation, close and reporting cycles by days or weeks..
- Facilitated conducting what-if scenarios for different revenue projections or changes in business lines.
- Facilitated tracking key corporate performance indicators from desktop.

Beside these impacts on the performance of the Sales function, it enabled giving brand, marketing, and sales managers the knowledge they need to strongly impact the top line through the use of brand, sales, promotion/marketing and delivering a full range of sales management analyses. With the IDB tool, near-real time measures by account, channel/channel segment, promotion, and campaign - can be used to improve the effectiveness of other business cycles such as the full range of brand, portfolio, and product analyses along with the ability to ask random, ad hoc questions and alert when actual performance varies substantially from plan.

In a nutshell, the IDB tool enabled getting the right information, to the right decision makers, at the right time. It is an enterprise-wide platform that supported multi-dimensionality reporting, analytics and decision modeling leading to fact-based decision making and getting a "single version of the truth."

The integration of the marketing data warehouse with the IDB tool also resulted in many benefits such as:

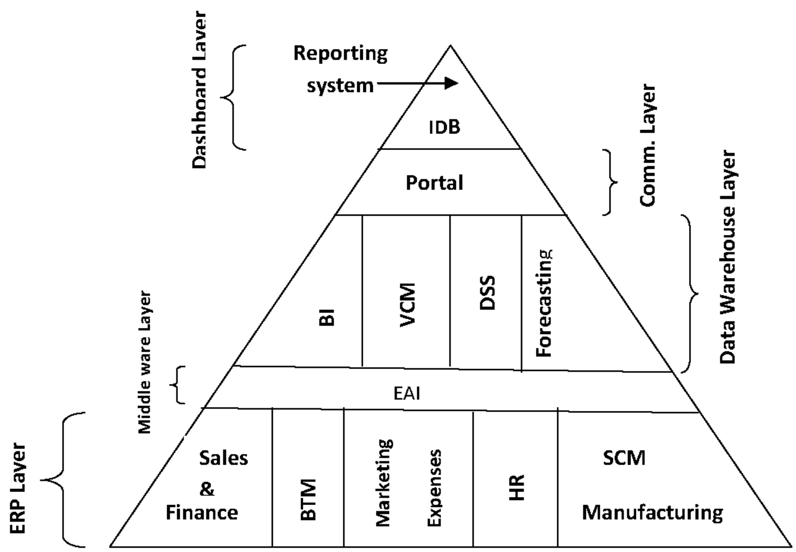
1. Financial process management of annual marketing budgets. Through the marketing Business Warehouse the brand managers gained autonomy in managing this process themselves through accessing this option on the IDB, where they would directly input their budgets online and the embedded workflow would proceed to request granting the necessary approvals. Budget managers were also able to perform internal budget shifts if need be as well as raising new purchase orders where previously it was done manually.
2. Ensuring internal system adherence to corporate financial and procurement procedures, there was no need for manual issuance any more as the ERP system provided the necessary constraints through budget coding, online approval requirements through a hierarchy workflow.
3. Monitoring marketing spend per brand and ensuring correct allocation across the marketing mix. The IDB reports were very malleable in their structure - can be driven by specific periods, campaign types, seasonality, segmentation by below the line marketing activities, or above the line, even by media type which allowed brand managers and marketing managers a better ability to assess the direct impact of a campaign spend over sales in a particular moment in time.
4. Providing budget reporting across franchise functions. Again, this has been a dissolved manual role as the new system allowed for automatic retrieval as explained in the point above.
5. Reporting on brand contribution and profitability per brand, pack, flavor and concentrate. The integration with the company-wide planning system allowed further report segmentation by account, for example, how each cost contributed to the overall profitability of a brand by integrating other operating expenses (OPEX), capital expenses (CAPEX) reports- this type of reports are high level top-line data that were accessed by top managerial staff and also depended on the level of data-breakdown details that were fed in by accountants.

The final Information Systems infrastructure architecture of the company after the IDB implementation is depicted in Figure 4.

CURRENT CHALLENGES FACING THE CASE FIRM

The ERP system has been put in place in the company for many years. Employees did not have an option, they had to use it. Though the company managed to streamline and align some

Figure 4. Architecture of the IS layers used in the company after IDB implementation



of its business processes and operations to the "best practice" processes embedded in the ERP, with minimum changes, to adapt to the specific nature of the beverage industry, customized bolt-in additional tools were inevitably used to cater for those specific characteristics. This was quite evident in the sales and marketing functions. As a result of the deployment of the new IDB system, several problems showed up.

The following obstacles and challenges were observed after installing the IDB

- User reluctance to abandon the use of old Excel templates increased resistance of the use of the new developed data entry facility.
- Training analysts to use IDB was a challenging feat. The dispersed remote locations of analysts made it difficult and costly to provide them with the necessary skills for optimum use of the new tool. Though conference calls and video-conferencing were used to tackle this obstacle, the level of operational excellence in using the tool was less than satisfactory as postulated by the management. Incompetency of use of the facilities of the system was also observed due to the lack of analytical proficiency.
- Unrealistic expectations were anticipated in a relatively short period.
- Lack of in-house technical expertise (The IDB support team is remotely located).
- Real-Time data reporting was accessible once accounting inputs are uploaded into the system in due course. Errors in data entry may thus lead to wrong information and decisions. As one employee contends:

"The initial step when we switched from the legacy system to the new setup involved data migration. This was a very tedious task and involved many months of intricate detailed work where existing data had to be assigned codes similar to the coding on the ERP to allow smooth transition. The challenge was to transition with almost zero error and entailed at certain moment of time to halt transactions entry to ensure all real-time data is

captured. There was a mechanism to double check on validity of data after transition."

- Employees were not receptive to adapt to change and resisted system implementation by initially rejecting the idea for months. Advanced training technology-based methods were adopted to allow for virtual help where global ERP Power Users were able to access local employees' screens to guide them through a step by step process. Adaptability took about one whole year as the users gained more confidence in system maneuver.
- Although the idea of the ERP is to eliminate a lot of paper work, the case was not so in Egypt in some areas as the embedded ERP workflow involved external as well as internal stakeholder's input in the system. For example, marketing assistants raising new purchase orders through the transaction modules were required to attach three different quotations from three different suppliers. The idea is that suppliers were to send those quotations through an interface mechanism that allowed automatic attach to the purchase order. Given that the business climate in Egypt is still underdeveloped in terms of technology maturity, many continued to send their quotations manually by courier messenger, or fax, rarely few at the time were used to the idea of sending through email. This created a bottle neck since the marketing assistant had to then scan those documents and upload them manually to process the purchase order – where a purchase order raising should have taken about an hour for completion ended up consuming a whole day while it has still not been sent to the brand manager for approval. In some situations this produced further time bottlenecks as some suppliers required down payments prior to activation of the required job on the system.
- Brand managers viewed their newly added task of managing the input-output process of budgets themselves and purchase order requisitions as a waste of time as they previously did not do so. They viewed it as a purely transactional administrative task that should be performed by the accounting department or rather assigned to a department secretary for coordination. In the beginning, they did try to perform the processing themselves, but later, decided that this administrative tasks have impeded their focus on managing the brand with consumers from a marketing perspective.
- An operational issue of ERP is currency variations. Although the system allowed for multiple currency input, the final marketing budget to be approved by division heads and headquarters had to be submitted and received in dollars. This created some issues where brand managers initially prepared their budget in Egyptian pounds per the various local quotations and pricing they received. This total amount however did not sometimes equate for the same amount assigned by global HQ on the ERP which was in dollars. The challenge for the finance was to constantly adjust rates according to fluctuations in exchange rates and other market factors such as inflation and so forth which also resulted in some chunks of the budget in Egyptian pounds being eaten away due to these adjustments. It continues to be an ordeal as concentrate pricing is also in US dollars while other transactions are in multiple currencies.
- The deployment of technology into everyday business is still not very mature within people's mindset in Egypt. For example, the lack of trust still persisted in the Egyptian ideology, where the deployment of online signatures through ERP

protocol was still viewed with a suspicious eye. Middle managers especially in the accounting unit still requested a pen-written on paper signature-it made them feel secure. However this required that paper formats of ERP documents were printed and managers were requested put their pen signatures. This did not help in achieving a more lean and streamlined operation which the system was supposed to create.
- End-User Support and Maintenance

A few years ago, for an employee to report an IT problem they had to call the local help desk and technical staff will attend to the call to solve the problem for him or her. However, the company HQ has developed a new process that employees experiencing problems will have to make a call to call center, give their employee ID and describe the technical problem about the software or the hardware. The call center, remotely accessed, will give the employee a ticket number with the problem code and send them an email summarizing the problem. The problem could be simple enough to be solved over the phone (i.e., the center tells the employee on the phone what he or she could do to solve the problem). If it's not then the center directs the problem to the local help desk to solve it. Then the center sends the employee an email asking if the problem has been solved or not to close the ticket. Some employees found this process difficult and time consuming

Future Measures to Enhance the Reporting Cycle and Recommendations

The preceding analysis revealed that key success measures that may help the company to overcome the obstacles cited and leads to turning the IDB system into a success story include the following:

- Executive management involvement and support
- Clear project ownership by including sales analysts in the project management team
- Proper planning
- Hard working and focus by all member countries' analysts
- Clear communication between business and IT by setting meetings, discussing problems and analyzing issues and providing continuous feedback
- Clear role definition for both analysts and IT professionals
- Instituting efficient and effective change management process
- Proper and continuous staff and analysts training

To that end, the company is planning to install new version of the IDB system in the upcoming years. The new system will allow users to create their own queries with specific design and more advanced sales breakdown. Also, the company is planning to use more advanced analytical tools such as the "Instant Visual Analysis" tool which is a simple versatile tool that will make volume analysis and graphic representation faster, easier and more flexible. The Instant Visual Analysis tool uses pivot table reports to help analyze and graph numerical data, answering questions, exhibiting trends, etc. With a few mouse clicks users can see who sold the most where, which brands were the most successful, and which pack sold best.

REFERENCES

Alter, S. (1980). *Decision support systems: Current practice and continuing challenge*. Reading: MA Addison Wiley.

Bancroft, N. (1996). *Implementing SAP R/3*. Greenwich, CT: Manning Publication Co.

Bingham, D. (1999). *Food and beverage companies need to integrate information enterprise-wide*. Beverage Online.

Brazel, J. F., & Dang, L. (2008). The effect of ERP system implementations on the management of earnings and earnings release dates. *Journal of Information Systems*, *22*(2), 1–21. doi:10.2308/jis.2008.22.2.1

Chatterjee, D., Grewal, R., & Sambamurthy, V. (2002). Shaping up for e-commerce: Institution enablers of the organizational assimilation of Web technologies. *Management Information Systems Quarterly*, *26*(2), 65. doi:10.2307/4132321

Chou, D., Tripuramully, H., & Chou, A. (2005). BI and ERP integration. *Information Management & Computer Security*, *13*(5), 340–349. doi:10.1108/09685220510627241

CRM2day. (2004). *Business intelligence*. Retrieved from www.crm2day.com/bi

Data monitor. (2001). *Business intelligence from data to profit*. Retrieved from www.researchandmarkets.com

Davenport, T. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*, *76*(4), 121–131.

DeLone, W., & McLean, E. (1992). Information systems success: The quest for the dependable variable. *Information Systems Research*, *3*(1), 60–95. doi:10.1287/isre.3.1.60

DeLone, W., & McLean, E. (2003). The DeLone and McLean Model of Information Systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30.

Ding, D., & Chen, J. (2007). Supply chain coordination with contracts game between complementary suppliers. *International Journal of Information Technology & Decision Making*, *6*(1), 163–175. doi:10.1142/S0219622007002332

Drucker, P. (1998). The next information revolution. *Forbes.* Retrieved from www.forbes.com

Fan, M., Stallaert, J., & Whinston, A. (2000). The adoption and design methodologies of component-based enterprise systems. *European Journal of Information Systems*, *9*, 25–35. doi:10.1057/palgrave.ejis.3000343

Francalanci, C. (2001). Predicting the implementation effort of ERP projects: Empirical evidences on SAP R/3. *Journal of Information Technology*, *16*(1), 33–48. doi:10.1080/02683960010035943

Fryling, M. (2005). ERP implementation dynamics. *Information Science and Policy.* University at Albany. State University of New York, Oct. 2005.

Gable, G., Sedera, D., & Chan, T. (2003). Enterprise systems success: A measurement model. *Proceedings of the 24th ICIS,* (pp. 576-591). Seattle, Washington.

Gable, G., van Den Heever, R., Erlank, S., & Scott, J. (2001). Large packaged application software maintenance: A research framework. *Journal of Software Maintenance and Evolution: Research and Practice*, *13*(6), 351–371. doi:10.1002/smr.237

Holsapple, C., & Whinston, A. B. (1996). *Decision support systems: A knowledge based approach*. Minneapolis, MN: West Publishing.

Ifinedo, P. (2006a). Extending the Gable et al. enterprise systems success measurement model: A preliminary study. *Journal of Information Technology Management, 17*(1), 14-33.

Ifinedo, P., & Nahar, N. (2007). ERP system success: An empirical analysis of how two organizational stakeholder groups prioritize and evaluate relevant measures. *Enterprise Information Systems*, *1*, 25–48. doi:10.1080/17517570601088539

Kallinikos, J. (2004). Deconstructing information packages: Organizational and behavioral implications of ERP systems. *Information Technology & People*, *17*, 8–30. doi:10.1108/09593840410522152

Kamaladevi, B. (2010). Customer experience management in retailing. *Business Intelligence Journal*, *3*(1), 37–54.

Krumbholz, M. (2000). Implementing enterprise resource planning packages in different corporate and national cultures. *Journal of Information Technology*, *15*(4), 267–280. doi:10.1080/02683960010008962

Kumar, K., & Van Hillegersberg, J. (2000). ERP: Experience and evolution. *Communications of the ACM*, *43*, 23–26.

Kumar, V., Maheshwari, B., & Kumar, U. (2001). An investigation of critical management issues in ERP implementation: Empirical evidences from Canadian organizations. *Technovation*, *23*(10).

Light, B. (2001). The maintenance implications of the customization of ERP software. *Journal of Software Maintenance and Evolution: Research and Practice*, *13*(6), 415–429. doi:10.1002/smr.240

Madapusi, A., & Kuo, C. (2007). Assessing data and information quality in ERP systems. *Proceedings of the Decision Sciences Institute Annual Meeting*, Arizona.

Madapusi, A., Kuo, C., & White, R. (2007). A critical factors approach to ERP information quality and decision quality. *Proceedings of the Decision Sciences Institute Annual Meeting,* Arizona.

Markus, M., & Robey, D. (1988). Information Technology and organizational change: Causal structure in theory and research. *Management Science*, *34*, 583–598. doi:10.1287/mnsc.34.5.583

McDonald, k., et al. (2002). *Mastering SAP business information warehouse*. Canada: Wiley Publishing.

Mensching, J., & Corbitt, G. (2004). EPR data archiving- a critical analysis. *Journal of Enterprise Information Management*, *17*(2), 131–141. doi:10.1108/17410390410518772

Nah, F. H., Faja, S., & Cata, T. (2001). Characteristics of ERP software maintenance: A multicause study. *Journal of Software Maintenance and Evolution: Research and Practice*, *13*(6), 339–414. doi:10.1002/smr.239

Nicolaou, A. (2004). *ERP system implementation drivers of post-implementation success*. Decision Support in an Uncertain and Complex World: The IFIP TC8/WG8.3 International Conference, 2004, (pp. 589-597).

Nicolaou, A., & Bhattacharya, S. (2007). Organizational performance effects of ERP systems usage: The impact of post-implementation changes. *International Journal of Accounting Information Systems*, *7*(1), 18–35. doi:10.1016/j.accinf.2005.12.002

O'Leary, D. (2002). *Enterprise resource planning systems: System lifecycle, electronic commerce, and risk*. Cambridge, UK: Cambridge University Press.

Olszak, C., & Ziemba, E. (2007). Approach to building and implementing business intelligence system. *Interdisciplinary Journal of Information, Knowledge, and Management*, *2*, 135–148.

Olszak, C., & Ziemba, E., (2006). Business intelligence systems in the holistic infrastructure development - supporting decision-making in organizations. *Interdisciplinary Journal of Information, Knowledge, and Management, 1*.

Parr, A., & Schanks, G. (2000). A model of ERP project implementation. *Journal of Information Technology*, *15*(4), 289–304. doi:10.1080/02683960010009051

Ptak, C., & Schragenheim, E. (2000). *ERP: Tools, techniques and applications for integrating the supply chain*. London, UK: Series on Resources Management, St Lucie Press/APICS.

Rasmussen, N., Goldy, P., & Solli, P. (2002). *Financial business intelligence: Trends, technology, software selection and implementation*. New York, NY: Wiley.

AMR Research (currently part of Gartner inc.). (2005). *Market Analytix Report: ERP 2004-2009*.

Robey, D., & Boudreau, M. (1999). Accounting for the contradictory organizational consequences of information technology. *Information Systems Research*, *10*, 167–185. doi:10.1287/isre.10.2.167

Seddon, P. B. (1997). A re-specification and extension of the DeLone and McLean model of IS success. *Information Systems Research*, *18*(3), 240–253. doi:10.1287/isre.8.3.240

Sedera, D., Gable, G., & Chan, T. (2004). Measuring enterprise systems success: The importance of a multiple stakeholder perspective. *Proceedings of the 12th European Conference on Information Systems*, (pp. 1-13). Turku, Finland.

Soh, C., & Tay-Yap, J. (2000). Cultural fits and misfits: Is ERP a universal solution? *Communications of the ACM*, *43*(4), 47–51. doi:10.1145/332051.332070

Spott, D. (2000). Componentizing the enterprise applications packages. *Communications of the ACM*, *43*(4), 63–90. doi:10.1145/332051.332074

Stedman, C. (1999, November 1). Failed ERP gamble haunts Hershey. *Computerworld*. Retrieved April 16, 2006, from www.computerworld.com

Stedman, C. (2002). Maximizing the ERP investment. *Competitive Financial Operations: The CFO Project*, *1*, 1–6.

Xu, H., Nord, J., Brown, N., & Nord, D. (2002). Data quality issues in implementing an ERP. *Industrial Management & Data Systems*, *102*(1), 47–58. doi:10.1108/02635570210414668

Yen, D. C., Chou, D. C., & Chang, J. (2002). A synergic analysis for Web-based enterprise resource planning system. *Computer Standards & Interfaces*, *24*(4), 337–346. doi:10.1016/S0920-5489(01)00105-2

KEY TERMS AND DEFINITIONS

Business Intelligence (BI): Refers to computer-based techniques used in analyzing business data, such as sales by products and/or departments or associated costs and incomes. In addition, BI technologies avail historical, current, and predictive views of business operations. Common functions of Business Intelligence technologies are reporting, online analytical processing, analytics, data mining, business performance management, benchmarking, text mining, and predictive analytics.

Dashboard: Is the application of visual iconic tools that indicate the status of a particular measurable quantity, event or value. Within the context of business performance measurement, the use of the dashboard may aide decision makers in quickly spotting the status of some KPIs of interest to them.

Data Mining: Is the process of extracting patterns from data. Data mining is becoming an increasingly important tool to transform the data into information. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery. Data mining is often used to uncover patterns in data pertaining to functional or departmental data sets.

Data Warehouse: Is a repository (collection of resources that can be accessed to retrieve information) of an organization's electronically stored data, designed to facilitate reporting and analysis. Data warehousing arises when an organization need reliable, consolidated, unique and integrated analysis and reporting of its data resources, at different levels of aggregation.

Decision Support Systems (DSS): Constitute a class of computer-based information systems including knowledge-based systems that support decision-making activities. DSSs serve the management, operations, and planning levels of an organization and help to make decisions, which may be rapidly changing and not easily specified in advance.

Enterprise Resources Planning (ERP): Is an integrated computer-based software system used to manage organizational internal and external resources. It is an architecture which facilitates the flow of information between all business functions within the boundaries of the organization and manages the interfaces with outside stakeholders.

Key Performance Indicators (KPI): Measures commonly used to help organizations define and evaluate how successful their businesses are, typically in terms of making progress towards long-term organizational goals. KPIs can be specified by answering the question, "What is really important to different organizational stakeholders?"

This work was previously published in Cases on Business and Management in the MENA Region: New Trends and Opportunities, edited by El-Khazindar Business Research and Case Center, pp. 156-176, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 59
Sharing Scientific and Social Knowledge in a Performance Oriented Industry:
An Evaluation Model

Haris Papoutsakis
Technological Education Institute of Crete, Greece

ABSTRACT

The chapter evaluates the contribution of shared knowledge and information technology to manufacturing performance. For this purpose, a theoretical model was built and tested in praxis through a research study among manufacturing, quality and R&D groups. The social character of science is perceived as a matter of the aggregation of individuals, not their interactions, and social knowledge as simply the additive outcome of mostly scientists, members of the three groups, making sound scientific judgments. The study results verify the significant contribution of shared knowledge to the manufacturing group performance. They also demonstrate that information technology influences notably the manufacturing group performance and, in a less significant way, the sharing of knowledge. Study results are useful to researchers and the business community alike as they may be used as a springboard for further empirical studies and can help put together strategies involving knowledge management and information technology.

INTRODUCTION

At the turn of the twentieth century many companies (BP, Canon, GlaxoSmithKline, Honda, Siemens and Xerox, among them) have tried, with varied achievement rates, to leverage knowledge assets by centralizing Knowledge Management (KM) functions or by investing heavily in Information Technology (IT) (Davenport and Prusak, 2000; Hansen and von Oetinger, 2001). In parallel, the number of new knowledge management articles, according to Despres and Chauvel (2000, p. 55) "... has more than doubled each year over

DOI: 10.4018/978-1-4666-1945-6.ch059

the past decade". Among them quite a few have proposed and tested models for the management of knowledge, with or without the support of information technologies (Knight, 1999; Larsen et al, 1999; Liebowitz et al, 2000; Kingsley, 2002). A considerably smaller number of such studies have investigated into how companies can leverage knowledge in order to improve business performance (Nelson and Cooprider, 1996; Chong et al, 2000; Firestone, 2001). Only one (Lee and Choi, 2003), among the articles reviewed for this study is combining all three variables: KM, IT and performance. This is exactly the gap this chapter is coming to fill in. Based on careful analysis of the above mentioned previous empirical studies, it builds and empirically tests a model that simultaneously explores the relationships among these three variables and their antecedents.

The chapter is organized in six sections. In the following section the theoretical framework is defined and a brief presentation of relevant previous empirical studies, focused on the links among knowledge management and information technology to business performance is given. In section three, we situate our own model within the above framework. The variables and the investigation hypotheses are defined. In section four, the research methodology is presented and details are given on the questionnaires –the principal research instruments– and the indicators used for construct measurement. In section five, the investigation hypotheses are tested, using regression analysis, and statistical data are given on questions not analyzed elsewhere. Finally, in section six, conclusions are summarized and recommendations are given for managers of collaborating groups in order to increase shared knowledge and to positively affect manufacturing performance.

THEORETICAL BACKGROUND

In the relevant literature, most attempts to investigate the links among KM and IT that lead to improved business performance, are done within the environment of the knowledge-creating company (Nonaka 1991; Nonaka and Takeuchi 1995). Building upon this pioneer work, Grant, in a series of articles (1995 with Baden-Fuller, 1996a, 1996b, 1997) and Sveiby (1997, 2001) presented in a very clear way the fundamentals of a knowledge-based theory of the firm. According to Grant (1997) –recapitulating on his previous work– the knowledge-based view is founded on a set of basic assumptions. First, knowledge is a vital source for value to be added to business products and services and a key to gaining strategic competitive advantage. Second, explicit and tacit knowledge vary on their transferability, which also depends upon the capacity of the recipient to accumulate knowledge. Third, tacit knowledge rests inside individuals who have a certain learning capacity. The depth of knowledge required for knowledge creation sometimes needs to be sacrificed to the width of knowledge that production applications require. Fourth, most knowledge, and especially explicit knowledge, when developed for a certain application, ought to be made available to additional applications, for reasons of economy of scale.

Theoretically, our research stands upon the 'knowledge-based theory of the firm' (Grant, 1997; Sveiby, 2001). The fundamental problem in traditional management theory is how to align the objectives of workers with those of managers and the stakeholders. In accordance with the knowledge-based view, "… if knowledge is the preeminent productive resource, and most knowledge is created by and stored within individuals, then employees are the primary stakeholders" (Grant 1997, p. 452). Under this perspective, management's principal challenge is to establish the mechanisms for collaborating individuals and groups to coordinate their activities in order to best integrate their knowledge into productive activity. Sveiby (2001) believes that people can use their competence to create value in two directions: by transferring and converting knowledge externally or internally for the organization they belong to. When the managers of a firm direct the efforts of their employees internally, they create tangible goods and intangible structures such as better processes and new designs for products. When they direct their attention outwards, in addition to delivery of goods and money they also create intangible structures, such as customer relationships, brand awareness, reputation and new experiences for the customers.

For Fukayama (1999), the existence of 'social capital' that serves as a glue to hold diverse constituencies together, is a primary cause of success or failure of any organization. The World Bank defines social capital as "norms and social relations imbedded in social structures that enable people to coordinate actions and achieve desired goals", a definition that applies to countries, societies or organizations. It is here where social knowledge has an important role to play. Individuals develop social knowledge through their interactions with the social environment. Stable systems of social knowledge are organized around certain domains; the collaborating groups in our study. According to Turiel (1983) the acquisition of social knowledge can be interpreted in two different ways: (i) it can be knowledge transmitted to the individual by other persons, and in this case the knowledge acquired is dependent on what is transmitted; or (ii) it can be knowledge constructed by individuals specifically about certain social phenomena (p.1). In an effort to capture the dialectic and dynamic relationship between the individual and social knowledge, Jovchelovitch (2007) develops a social-psychological approach in order to investigate knowledge in every day life. In her framework, problems of social knowledge are discussed in relation to individual, social and collective representations. Knowledge represents at the same time subjective, inter-subjective and objective worlds (p. 168).

It is under the above theoretical perspective that we are reviewing the literature, relevant to our investigation, in the following section.

Previous Empirical Studies

Linking knowledge management and information technologies with business performance has never been an easy task. Comparing KM projects to their two prevailing predecessors (total quality management and business process re-engineering) Armistead (1999) notices that authors on KM "… do not use the same hard measures of success consistently" (p. 143). He believes that for a knowledge-based view to be useful, it must help improve some key performance indicators like quality, flexibility and cost. Referring to manufacturing companies he notes that operational processes, which depend more on knowledge, are expected to perform well against measurements of quality in consistency, while at the same time they improve productivity.

Our research focused on two basically diverse areas: The measurement –in terms of both qualitative and quantitative results– of a KM project's impact and, at the same time, the identification of the cause-effect relationship that exists between KM, IT, and the overall business performance. Some previous studies captured KM contribution by focusing on intellectual capital measures (Larsen et al, 1999) or accounts and audits (Liebowitz et al, 2000) but both groups of authors question the generability of their studies. Other studies, criticizing conventional performance measures –such as Return On Investment (used by Anderson, 2002) and Economic Value Added, used by multinationals like The Coca-Cola Company– propose measures based on the Balanced Scorecard (Knight, 1999) or other more abstract and tailored to the company, like the Comprehensive Benefit Estimation (Firestone, 2001) and the Cost of Information (Kingsley, 2002). In a recent work the relevant literature summarized above has been extensively reviewed (Papoutsakis and Salvador Valles, 2006).

In most of the above empirical studies the role of shared knowledge among company departments is not consistent, despite the fact that the knowledge transfer process has been studied extensively. Trust and influence have only been recognized as antecedents of shared knowledge by Nelson and Cooprider (1996), while Lee and Choi (2003) consider trust and information technology as knowledge creation enablers among seven others. What is really missing is an integrative model combining shared knowledge and information technology with performance. Although several studies investigate the relationship between KM and performance (Nelson and Cooprider, 1996; Chong et al, 2000; Firestone, 2001) or IT and KM (Lee and Choi, 2003), they fail to explore the relationships among KM, IT and performance simultaneously.

It is believed that if managers become conscious of the fact that these relationships have interactive features, they can stand a much better chance of improving the performance of their department or company. Measuring the impact of shared knowledge and IT upon manufacturing performance is not an easy task as this will strongly affect the behaviour of managers and employees not only of the manufacturing group, but those of the collaborating groups (in our case the quality and R&D groups). Regarding social knowledge, we have to consider that it exists in the relationships, not in the individuals themselves and thus it requires mutual commitment, since if one party withdraws it disappears. It is under this perspective that we have built and empirically tested the evaluation model proposed in the following section.

PROPOSED MODEL

Aiming to gain insight into the essential factors influencing manufacturing performance, the development and testing of a conceptual model containing the minimum selected theoretical constructs, is considered. Three have been our major concerns, upon building our research model. First, we did not want to propose a model that delineates every possible variable or process that affects manufacturing performance. Second, we wanted to focus on shared knowledge as the leading expression of knowledge management, among the manufacturing, quality and R&D groups of a firm. Third, information technology, in our model, has been perceived to affect both manufacturing performance and shared knowledge.

To assess the type of knowledge to be shared was also an interesting question. Von Krogh, Ichijo & Nonaka (2000) define knowledge as a justified true belief: when somebody creates knowledge, he or she makes sense out of a new situation by holding justified beliefs and committing to them. The emphasis in this definition is on the conscious act of creating meaning. In our study, we focused on collective knowledge that entails notions of collective belief, truth and justification (Corlett, 1996). Our analysis insisted on particular conditions of inter-group, justified true acceptance which is necessary for collective knowledge. According to Corlett, "… what makes belief, acceptance, justification and knowledge collective is that they are the results of human decision-makers related to one another in groups…" (2007, p.245). Obviously, each one represents his or her group interests.

The road to sharing knowledge lies through individuals, mostly scientists in our study, and is based upon building social relationships and trust, deep dialogue and creative abrasion. There is a need of diversity of ideas and an environment where failures and reflection are valued as learning enablers. Science is the process used everyday to logically complete thoughts through inference of facts determined by calculated experiments. As science itself has developed, the so produced scientific knowledge has developed a broader usage within scientists. The development of scientific methods has made a significant contribution to our understanding of scientific knowledge. To be termed scientific, a method of inquiry must be based on the collection of data through observation and experimentation, and the formulation and testing of hypotheses.

The social dimension of scientific knowledge is of significant importance, as well. We perceive the social character of science as a matter of the aggregation of individuals, not their interactions, and social knowledge as simply the additive outcome of mostly scientists, members of the three groups, making sound scientific judgments. Philosophers concerned to defend the social character of knowledge and to explore the social dimension of scientific practice (Laudan, 1984; Brown 1989; Goldman, 1995) have approaches that differ in their details but they agree in stating that scientists are persuaded by what they regard as the best evidence or argument, the evidence most indicative of the truth by their lights, and in maintaining that arguments and evidence are the appropriate focus of attention for understanding the production of scientific knowledge. Opposing them, Jovchelovitch (2007) criticizes the narrow association of knowledge with rationalism in the sense of scientific knowledge. As a result, scientific knowledge is viewed as more valid than everyday knowledge.

Therefore, we have opted for our model to highlight a few key factors that can explain a large proportion of the variation noted in manufacturing performance. We have modified the sharing knowledge model validated and used by Nelson & Cooprider (1996) and we enhanced it with links allowing us to draw conclusions on the role and contribution of information technology as an enabler and facilitator towards both manufacturing performance and shared knowledge. Thus, the proposed evaluation model is built to investigate cause and effect links between sharing knowledge, its components, information technology and manufacturing performance.

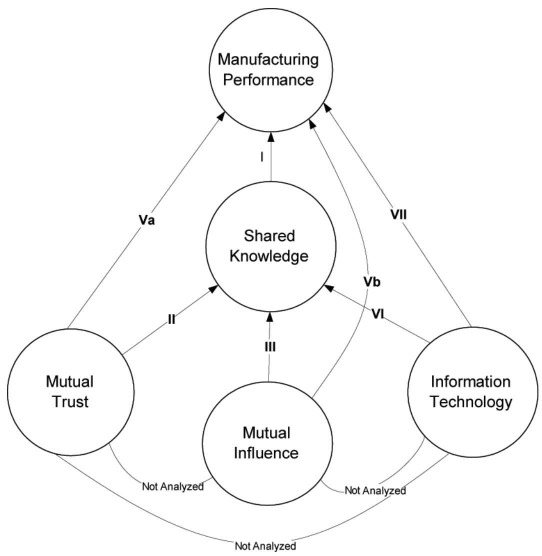
Both general and multiplicative methods are used to measure the indicators, at least two for every construct, and path analysis has been chosen as the analytic technique in this study because it assesses causal relationships (Pedhazur, 1982; Wright, 1971). Pedhazur, building upon Wright, states that "…path analysis is not a method for discovering causes, but a method applied to causal models formulated by the researcher on the basis of knowledge and theoretical considerations." (p. 580). Path diagrams, although not essential for numerical analysis, are useful tools for displaying graphically the pattern of causal relations among the set of variables under consideration. In this respect we consider the model more appropriate than the intellectual capital or the tangible and intangible approach used in other studies.

Despite the fact that in recent years, social and behavioural scientists have been showing a steadily growing interest in studying patterns of causation among variables, the concept of causation has generated a great deal of controversy among both philosophers and scientists. Nonetheless, causal thinking plays a very important role in scientific research. Even in the works of those scientists who strongly deny the use of the term causation, it is very common to encounter the use of terms that indicate or imply causal thinking. Thus, we can conclude that scientists, in general, seem to have a need to resort to causal frameworks, even though on philosophical grounds they may have reservations about the concept of causation.

Schematically, our empirical evaluation model illustrates the relationships among the five variables as shown in Figure 1. Our seven hypotheses correspond to the causal links of Figure 1 and derive from theoretical statements found in the literature related to knowledge management and information systems and technology. In the following section, we shall elaborate upon the variables incorporated in our model and, at the same time, we shall present our investigation hypotheses.

Finally, it is important to bear in mind that path analysis is a method, and as such its valid application is subject to the competency of the researcher using it and the soundness of the theory that is being tested. Finally, it is the explanatory scheme of the researcher that determines the

Figure 1. The shared knowledge and information technology evaluation model



type of analysis to be applied to data, and not the other way around.

VARIABLES AND HYPOTHESES

Shared Knowledge

Sharing of knowledge is a process distinct from managerial communication, which also deserves consideration. Nelson & Cooprider (1996, p. 411) define Shared Knowledge as "an understanding and appreciation among groups and their managers, for the technologies and processes that affect their mutual performance". Appreciation and understanding are the two core elements of shared knowledge. Appreciation among diverse groups must be characterized by sensitivity to the point of view and interpretation of the other group, in order to overcome the barriers caused by the different environments and languages used. A deeper level of knowledge must be shared in order to achieve mutual understanding and this is often characterized as organizational knowledge Badaracco (1991).

Lack of this organizational and cross-functionally shared knowledge may result in loses of Manufacturing group performance, while its presence may lead to better performance. As we do not have *a priori* reasons to expect a different relationship, it is here that we are founding our first hypothesis.

Hypothesis 1: *Shared knowledge among Manufacturing, R&D and Quality groups, as perceived by the manufacturing organization, leads to improved manufacturing group performance.*

In an effort to make more comprehensible the relationship between shared knowledge and the manufacturing group performance, we shall now define the two components or antecedents of shared knowledge: Trust and Influence.

Trust

The significance of trust has been given considerable attention and has even been described as a 'business imperative' (Davidow and Malone, 1992; Drucker, 1993 among others). In rather similar ways, trust has been defined as "a set of expectations shared by all those in an exchange" (Zucker, 1986) or as "the expectation shared by the [involved] groups that they will meet their commitments to each other" (Nelson and Cooprider, 1996, p. 413) or finally as "… maintaining reciprocal faith in each other in terms of intention and behaviors" (Lee and Choi, 2003, p. 190).

Szulanski (1996) empirically found that the lack of trust among employees is one of the key barriers against knowledge sharing and that the increase in knowledge sharing brought on by mutual trust results in knowledge creation. In the model proposed for this study, it is assumed that Manufacturing, R&D and Quality groups work better in an atmosphere of mutual trust based on mutual commitment and a stable long-term relationship, which is the foundation for our conceptualization of trust. We, thus, hypothesize that mutual trust is a determinant of shared knowledge and it is here that we advance our second hypothesis.

Hypothesis 2: *The perception of increased levels of mutual trust among Manufacturing, R&D and Quality groups leads to increased levels of shared knowledge among these groups.*

Influence

As organizational groups engaged in joint work are often dependent upon each other, influence relationships are created. One way influence is developed, is through the law of reciprocity (Cohen and Bradford, 1989). People expect payback for contribution to an exchange. The perception of reciprocal benefits leads to mutual influence and success in future exchanges among the groups. Nelson and Cooprider (1996, p. 414) define mutual influence as "the ability of groups to affect the key policies and decisions of each other." Consequently, we expect the following relationship to hold true and it is here that we are basing our third hypothesis.

Hypothesis 3: *Increased levels of mutual influence among manufacturing, R&D and Quality groups lead to increased levels of shared knowledge among these groups.*

The two important aspects with regard to shared knowledge are demonstrated in the evaluation model used for this research (Figure 1). First, mutual trust and influence are presented as antecedents of shared knowledge, and second, shared knowledge is presented as a mediating variable between mutual trust and influence, leading to manufacturing group performance. Therefore, we can hypothesize:

Hypothesis 4: *Shared knowledge acts as a mediating variable between mutual trust and influence and manufacturing performance.*

As we have no *a priori* reasons to exclude that mutual trust and influence could also possibly affect manufacturing performance directly, we are here introducing our fifth hypothesis.

Hypothesis 5: *There is a positive relationship between mutual trust and manufacturing performance, as well as between mutual influence and manufacturing performance.*

Information Technology

Davenport & Short (1990, p. 11) define Information Technology (IT) as "…the capabilities offered by computers, software applications, and telecommunications" and further explain that "IT should be viewed as more than an automating or mechanizing force; it can fundamentally reshape the way business is done" (p. 12) and that "IT can make it possible for employees scattered around the world to work as a team" (p. 19). Applegate, McFarlan & McKenney (1999; p. vii) identify IT as: "…computing, communications, business solutions and services…" and further down (note in p. 3) they explain that "…IT refers to technologies of computers and telecommunications (including data, voice, graphics, and full motion video)."

In the new economy era, information technology has a very significant role to play in supporting both communication and, in particular, knowledge sharing. IT affects knowledge sharing in a variety of ways. IT facilitates rapid collection, storage, and exchange of knowledge in a scale not possible up to recent times, thus fully supporting the knowledge sharing process (Roberts, 2000). Specially developed IT integrates fragmented flows of knowledge, eliminating, in this way, barriers to communication among departments (Gold et al, 2001). Advanced IT (like electronic whiteboarding and videoconferencing) encourages all forms of knowledge sharing and is not limited to the transfer of explicit knowledge only (Riggins and Rhee, 1999). Thus, we can hypothesize:

Hypothesis 6: *There is a positive relationship between IT support and the knowledge sharing process.*

The use of certain IT infrastructure such as intranets, extranets, groupware, internet, etc has been evaluated, in relationship to sharing knowledge, by means of an ad hoc question. IT, in our model, is perceived to affect manufacturing performance, as well.

Manufacturing Performance

Under an industrial business management approach, manufacturing performance has three main activities: (i) the selection of goals; (ii) the consolidation of measurement information relevant to an organization's progress against these goals, and (iii) the interventions made by managers in light of this information with a view to improving future performance against these goals. Although presented here sequentially, typically all three activities will run concurrently, with the interventions made by managers affecting the choice of goals, the measurement information monitored, and the activities being undertaken within the organization.

For the purpose of our study, organizational stakeholders in every participating company have been questioned in order to assess the manufacturing group performance and, in addition, to compare the manufacturing unit under investigation with other units they have managed. Madnick (1991) points out the major ways in which IT support affects manufacturing group performance. First, IT provides opportunities for increased inter- and intra-organizational connectivity and, thus, increases both efficiency and effectiveness. Second, new IT architectures offer significant cost/performance and capacity advances. And finally, with IT support, adaptable organizational structures that lead to significant cost reductions are made possible. As there are other variables (such as employees' competences and qualification, raw material quality, technology level of the machinery in use, etc) which affect manufacturing group performance and are not included in our model, we can only hypothesize:

Hypothesis 7: *There is a positive relationship between IT support and the manufacturing group performance.*

The use of four IT functions (coordination of business tasks, support of decision making, facilitating teamwork and access to information in data bases) has been evaluated, in relationship to manufacturing performance, by means of an ad hoc question.

The five variables incorporated in our model are structured upon a socio-technical perspective that adopts an holistic approach (Pan and Scarbrough 1998). Based on this view, mutual trust and influence, related to the organizational structure and culture as well as to the employees, are considered social variables while, on the other hand, IT is considered a technical variable. For purposes of clarity, most studies consider the impact of social and technical variables independently, a precaution we are also adopting in this study. In the next section, we are presenting the methodology of our research.

RESEARCH DESIGN

In an ideal situation, investigation samples are selected randomly. This is done, among other reasons, for the external validity criteria to be *a priori* fulfilled. The maxim applies to the selection of companies, manufacturing units, and, to a certain extent, to the selection of individuals who answer the questionnaires. As the sample of our study included every company that has accepted to participate we can not disregard a possible selection bias. Finally 51 medium to large size industrial companies, representing 5 sectors (alimentation, automotive, chemical and pharmaceutical, electro-mechanical, and textile) participated in the research. The unit of analysis is the manufacturing group, since the intent of this study is to explain the relationship of organizational subunits (the three collaborating groups) rather than that of individuals. The size of the company has been used as a criterion and it was convenient that several of the selected companies had multiple manufacturing groups (or departments/lines as they were named) who cooperated with a central R&D and/or quality group.

This has allowed for the research to be addressed to a big number of manufacturing groups, out of which 112 have participated by responding to the relevant questionnaires. Table 1 shows the industrial sectors represented, the number of companies contacted and participated as well as the identified and participating manufacturing units for each one of them. The final sample size, of 112 manufacturing units, is considered sufficient in order to perform path analysis (Pedhazur 1982) and the participation rates achieved in our study (62% at company level and 68% at the unit of analysis level) are considered satisfactory (Cook and Campbell, 1979).

The research responders have been chosen based on the key-informant methodology developed by Phillips and Bagozzi (1986) and included–for each company–manufacturing, R&D and quality group managers or their deputies, as well as senior managers. As the measurement of organizational characteristics requires research methods different from those used for measuring the characteristics of individuals, key-informant methodology is a frequently adopted approach. (Table 1)

Two symmetrical relationship questionnaires, worded in a reverse form, were addressed to Production and Quality or R&D managers -and their assistants- and aimed at portraying the opinion and the attitude of the two collaborating groups towards each other, in terms of sharing knowledge. In addition, the role and level of contribution of Information Technology, both as a tool and/or enabler in supporting sharing knowledge among the collaborating groups was investigated and a last, ad hoc question evaluated the use of commonly used IT infrastructure for inter-firm knowledge sharing.

A third, performance questionnaire –attempting to measure manufacturing group performance– was addressed to senior managers or their assistants. They have been asked to compare the manufacturing group under question, to other comparable manufacturing groups they have managed. In addition, the level of contribution of Information Technology to manufacturing group performance was investigated and again, a last ad hoc question evaluated the use of specific IT functions on four knowledge sharing issues, closely

Table 1. Study participants by sector, company and unit of analysis

Sector	Companies		Manufacturing Units	
	Contacted	Participated	Identified	Participated
Alimentation	26	14	47	31
Automotive	8	6	25	15
Chemical & Pharmaceutical	7	5	22	19
Electro-Mechanical	25	18	54	35
Textile	16	8	17	12
Total	82	51 (62%)	165	112 (68%)

related to the group performance. The questions used, with their indicative numbers, are listed in Appendix I, where we analyze the indicators used for each construct measurement.

The two relationship questionnaires were pilot tested using Production and Quality or R&D managers, and the performance questionnaire was tested using senior executives from a small group of companies not participating in the final phase of our research. Following the completion of each pilot questionnaire, the pilot test informant was debriefed to determine if any questions were confusing for any reason. They were also questioned, whether in their opinion, any significant indicators have been left out of the questionnaire. Based on the results of the pilot test, a number of initially used questions were determined to be poor and were deleted or rephrased. The most important lessons learned through design and pilot testing of the questionnaires are:

1. In designing the questions, it is essential, to word them in as simple terms as possible and to anchor each question to one specific relationship;
2. Each question must be customized to include the exact name of the department, as it is used in the company in question.

Despite the above precautions we experienced that the key-informant does not always share the same understanding with the researcher regarding the terminology in use.

Two types of measures have been used to assess the organizational characteristics of shared knowledge, mutual trust, mutual influence, information technology and manufacturing performance. General measures, where each informant is asked to assess the overall level of interaction for a specific characteristic of a particular relationship and multiplicative or interaction measures, where each informant is asked, for example, to assess the role of manufacturing and either R&D or quality group for each characteristic separately. Using the conceptualization of fit as interaction, proposed by Venkatraman (1989), the measurements have been operationalized as "manufacturing role X R&D or quality role", by multiplying the two responses together.

There are a number of advantages to this measurement scheme, as indicated by Churchill (1979) and Campbell and Fiske (1959): (a) the two types of measures (general and multiplicative) can be thought of as different methods; (b) it provides a stronger test of the validity of the measurement scheme, and (c) it balances possible threats to validity inherent in either type alone.

Manufacturing group performance has been conceptualized in two parts; as operational and service manufacturing performance. Operational or 'inward' performance is operationalized as: (a) the quality of the manufacturing group's work product; (b) the ability of the manufacturing group to meet its organizational commitment, and (c) the ability of the manufacturing organization to meet its goals (first three questions of the performance questionnaire). Service or 'outward' performance is operationalized as: (a) the ability of the manufacturing group to react quickly to R&D and/or quality needs, (b) its responsiveness to the R&D and/or quality group and (c) the contribution that the manufacturing group has made to the R&D and/or quality group's success in meeting its strategic goals (questions four to six of the performance questionnaire).

ANALYSIS OF THE RESULTS

In order to assess the validity of our evaluation model (Figure 1) we empirically tested it using path analysis as the method for studying patterns of causation within the set of independent, mediating and dependent variables used in our evaluation model. For the casual model under consideration, the following preconditions, given by Pedhazur (1982) are essential:

1. The relations among the variables in the model are linear, additive and causal.
2. Each residual is not correlated with the variables that precede it in the model.

This implies that:

a. The residuals are not correlated among themselves
b. All relevant variables are included in the model
c. Each endogenous variable is perceived as linear combination of exogenous and/or endogenous variables in the model plus a residual
d. Exogenous variables are treated as ‘given’ and when are correlated among themselves, these correlations are also treated as ‘given’ and remain unanalyzed.

3. There is a one-way causal flow in the system.
4. The variables are measured on an interval scale.
5. The variables are measured without error.

And Pedhazur concludes that “…given the above assumptions, the method of path analysis reduces to the solution of one or more multiple linear regression analyses” (p. 580).

It is under these assumptions that we have concluded to the use of Figure 2, as the research model for our investigation. With one exception: Not all variables affecting Manufacturing Performance are included in the model. Essential variables like skills and qualification of workers, technological level of the machinery in use, and quality of the raw material –just to mention some very basic ones- have not been taken into consideration simply because they do not relate to the focus of our investigation, which is the contribution of shared knowledge and information technology to manufacturing performance. This means that the result of the regression of Manufacturing Performance versus Shared Knowledge could only be considered as a partial causal effect.

Two multiple regressions were run for each of the two dependent variables, manufacturing performance and shared knowledge. Testing the hypotheses requires testing the significance of paths I, II, III, Va, Vb, VI and VII as presented in

Figure 2. Regressions in the evaluation model

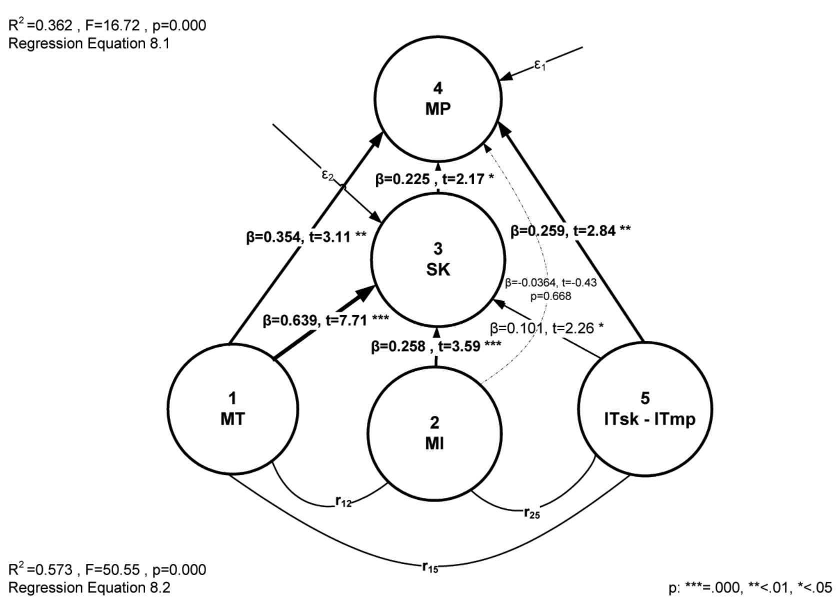


Figure 1. The results of this analysis are schematically shown in Figure 2 and in the generic regression equations below:

For manufacturing performance:

$$MPC = \alpha + \beta_1\ SKC + \beta_2\ MTC + \beta_3\ MIC + \beta_4\ ITmpC + e \quad (8.1)$$

For shared knowledge:

$$SKC = \alpha + \beta_1\ MTC + \beta_2\ MIC + \beta_3\ ITskC + e \quad (8.2)$$

Two points need to be clarified in the above equations:

1. β's, the normalized path coefficients, indicate the direct impact of a variable hypothesized as a cause on a variable taken as an effect. Wright (1934) defines a path coefficient as: "The fraction of the standard deviation of the dependent variable (with the appropriate sign) for which the designated factor [here, the independent or mediating variable] is directly responsible…" (p. 162). Under the previously analyzed preconditions, path coefficients take the form of ordinary least squares solutions for the β's (Pedhazur 1982, pp. 582-584).
2. The third letter C, added to the two-letter acronym used for each one of the variables, indicates that we are referring to its Construct. As at least two indicators have been used to assess every variable in the research model, the construct is the mean of these indicators. In the acronym of information technology, the indicators mp and sk are used to distinguish: (a) ITskC, the IT construct measured through the two relationship questionnaires, in reference to shared knowledge, and (b) ITmpC, the IT construct measured through the performance questionnaire, in reference to manufacturing performance. As these two types of questionnaires have been filled in by different key-informants we could not use a possible IT Construct (ITC) produced as the mean of ITskC and ITmpC.

Regressions in the evaluation model have been conducted in hierarchical order. First, we examined the relationship between manufacturing performance and each one of the variables affecting it; shared knowledge, mutual trust and influence, and information technology as described in the first regression equation. And the resulting equation is:

$$MPC = 6.98 + 0.354\ MTC - 0.0364\ MIC + 0.225\ SKC + 0.259\ ITmpC + e_1$$

At this point, and for the better understanding of the analysis following, some more statistical terms need to be clarified:

1. R^2, in the case of multiple independent variables, indicates the squared multiple correlation, i.e. the proportion of variance of the dependent variable accounted for by the independent variables.
2. The t-value ($-\infty < t > +.\infty$) determines the level of significance of the β's, and finally,
3. F, the ratio of the mean square regression to the mean square residual, provides a statistic for testing the null hypothesis. When the calculated F exceeds the tabled value of F, with the associated degrees of freedom and at a preselected level of significance p (i.e. $p=0.000$, or $p<.05$), the conclusion is to reject the null hypothesis.

In this first regression mutual trust, information technologies and shared knowledge are found to affect manufacturing performance significantly, while mutual influence does not (β=-0,0364, t=-0.43, p=0.668). The regression model described by the equation 5.1 is significant (F=16.72, p=0.000), but R^2=0.362 suggests that only 36.2 percent of the variance is explained by the five variables involved. This is something we expected, as we have already noted, upon founding hypothesis 7, that there are significant factors affecting manufacturing performance which are not included in our model.

Then, we examined the relationship among shared knowledge, mutual trust and influence, and information technology, as described by the second regression equation, and here are the results:

$$SKC = 1.08 + 0.639\ MTC + 0.258\ MIC + 0.101\ ITskC + e_2$$

In this second regression, mutual trust, mutual influence and information technology are all found to affect shared knowledge with variable strengths. The regression model described by the equation 5.2 is significant (F=50.55, p=0.000) and R^2=0.573 suggests that 57.3 percent of the variance is explained by these three variables.

Consistency of the model with the data, however, does not constitute proof of the theory; at best it only provides support to it. Following Popper's (1959) basic argument that all one can achieve through investigation is the falsification of theory, we would have to conclude that the theory has survived the test, in that it has not been disconfirmed. Thus, in direct connection with our investigation hypotheses, the regression results indicate that:

1. Hypotheses 1, 2, 3, 6 and 7 are directly supported by the significance of paths I, II, III, VI and VII respectively. This means that: (a) Shared knowledge among Manufacturing, R&D and Quality groups, as perceived by the manufacturing organization, leads to improved manufacturing group performance. (b) The perception of increased levels of mutual trust among Manufacturing, R&D and Quality groups leads to increased levels of shared knowledge among these groups. (c) Increased levels of mutual influence among Manufacturing, R&D and Quality groups lead to increased levels of shared knowledge among these groups. (d) There is a positive relationship between IT support and the Knowledge sharing process. (e) There is a positive relationship between IT support and the manufacturing group performance.
2. Hypotheses 4 and 5 are each only partially supported as they could not both together stand true. This means that: (a) Shared knowledge acts as a mediating variable only for mutual influence, while mutual trust appears to also significantly affect manufacturing performance in a direct way (significance of path Va). (b) Mutual influence does not directly affect manufacturing performance (statistically insignificant *beta* for path Vb).

There is an important note to be made at this point. To the extent that *beta* values reflect the strength of the cause-effect relationship, we may say that IT does not affect shared knowledge in the same significant way that it affects manufacturing performance. This result may first be explained by the fact that information technologies mainly affect transfer and sharing of explicit knowledge, while in the environment of our study (shared knowledge among manufacturing, quality and R&D groups) tacit knowledge plays a dominant role. The result is also in accordance with findings of other studies. Lee and Choi (2003) have found that IT support is significantly related only with knowledge combination (explicit to explicit knowledge transactions) while they have noticed no significant relation with any of the other three knowledge creation processes (socialization, externalization and internalization) where tacit knowledge is also involved.

The second explanation has to do with the research instruments. In our investigation, the two constructs of information technology (ITskC and ITmpC) were measured on two separate instruments, the symmetrical relationship questionnaires, and the performance questionnaire. The two separate instruments were filled out by different key-informants at different levels within the organization. It is anticipated that collaborating group managers on one hand, and senior managers on the other, might have different background conditions, when asked to judge the same concept. Pedhazur (1982) attributes these differences to the personal characteristics of key-informers, like cognitive styles, self-concept, ego strength and attitudes.

Use of IT Infrastructure

The last question in the two relationship questionnaires is examining the use of certain IT infrastructure as tools and enablers for sharing knowledge, among Manufacturing, Quality and/or R&D groups. Study results indicate that managers or their deputies of the three collaborating groups strongly use E-mail (86.6%), Intranets (71%) and Internet (42.85%), and at lower, but still noteworthy percentages, Data Warehouse software (30%), Extranets (23.65%), Groupware software (20.95) and Workflow software (11.6%) in their daily work. Percentages here and in the following paragraph refer to the sum of 'strong' answers (grouped Likert ratings 5, 6 and 7).

Use of IT Functions

The last question in the manufacturing performance questionnaire is investigating the use of certain IT functions by the company as a whole. According to our study, senior managers report that group managers use at relatively high percentages all four IT functions, in order to: facilitate access of information in Data Bases (84.4%), coordinate business tasks (82.6%), facilitate team members to work together (76.4%) and support decisions making (69.2%).

Confirmatory Tests

Four confirmatory tests of the research model were performed and the results obtained are briefly presented here. Cronbach's *alphas* (all ranged from 0.7819 to 0.9994) were utilized to reassure the reliability of the instruments used (Nunnally, 1978).

Convergent and discriminant validity has been checked by the Multi-Trait Multi-Method correlation matrix and all correlations within constructs have been found to be higher than any correlations across constructs (Campbell and Fiske, 1959).

Linearity and collinearity tests are essential for the assumptions of regression analysis to be met. Because the scatter plots of individual variables did not indicate any nonlinear relationships, the linearity was guaranteed. In addition, we tested the plots of residuals against the explanatory variables. As they showed no model inadequacies, we assume that no variable violates the constant variance. Collinearity among the variables involved in the two regression equations was tested by the Variance Inflation Factors which in our study ranged from 1.1 to 2.3 (in the first regression equation) and from 1.1 to 1.3 in the second. Hence, we have taken for granted that there is no multicollinearity problem (Neter et al, 1996).

Finally, the analysis of variance was used to check via an alternative method, two of the results obtained through Multiple Regression (Pedhazur, 1982; Draper and Smith, 1980).

1. The corresponding values of r (or $R\text{-}S_q$) were re-calculated and found in accordance with the previously calculated R's:
   r= 0.3846 compared to R= 0.362 for the first regression equation, and
   r=0.58406 compared to R= 0. 573 for the second regression equation.
2. The regression models used were found significant because both F-ratios were larger than the corresponding critical F-values:
   F=16.72 >> F(0.01; 4, 107) = 3.50 for the first regression equation, and
   F=50.55 >> F(0.01; 3, 108) = 3.96 for the second regression equation.

The statistical results of the two regression equations and those of the confirmatory tests are presented in Appendix II.

LIMITATIONS AND FUTURE RESEARCH

We acknowledge two limitations for our study. The first one is theoretical:

1. The development of mutual trust and influence leading to shared knowledge and the influence of information technology are all ongoing phenomena. In our study, these constructs were measured at a static point in time rather than as they develop. A future research could possibly investigate the relationship of ongoing changes to manufacturing group performance, maintaining the same company sample. It would also be interesting to possibly relate the changes noted over time, with actual changes in both the social (mutual trust and influence) and the technical (information technology) subsystems within the organization.
2. The study was conducted in Spain. A new multinational study in three more European Union countries, namely Finland, Greece and Hungary is currently under development and we hope that it will further support our findings.

CONCLUSION

The results of this study demonstrate the positive contribution of shared knowledge and information technology to manufacturing performance. Based primarily on the above results and to a certain extent on the literature reviewed, we come to the following socio-technical conclusion. Sharing knowledge in a meaningful manner requires a well balanced merge of technology with the company's culture, in a way that creates an environment supporting collaboration. Trust has been identified, through our study, as one of the company's core values. Management has to create a climate of trust in the organization, for knowledge sharing to become reality. In such an environment scientists from different groups (Manufacturing, Quality and R&D) feel comfortable to look for others with the 'missing piece of individual and social knowledge' to share. As shown by this study, influence is the second necessary condition for, and can lead to cooperative behavior among individuals and groups, especially where tacit knowledge has to be shared. It is only in such an environment that the IT made available may lead to innovative products.

The findings of this study indicate that Manufacturing, Quality and R&D groups have the opportunity to develop mutual trust and influence through repeated periods of positive face-to-face or IT-based communication, social interaction and common goal accomplishment. Such behavioral features result to increased shared knowledge regarding the groups' common problems, procedures and know-how. It is clearly illustrated that it is in the hands of management to increase manufacturing performance by improving the channels for individual and social knowledge to be shared among the three groups and by selecting the information technologies that best fit the innovative efforts and competitive strategy of their organization.

REFERENCES

Anderson, M. (2002). Measuring Intangible Value: The ROI of Knowledge Management. Retrieved July 12, 2004, from http://www1.astd.org/news_letter/november/Links/anderson.html

Applegate, L. M., McFarlan, F. W., & McKenney, J. L. (1999). *Corporate Information Systems Management: Text and Cases* (5th ed.). USA: Irwin/McGraw-Hill.

Armistead, C. (1999). Knowledge Management and Process Performance. *Journal of Knowledge Management*, *3*(2), 143–154. doi:10.1108/13673279910275602

Badaracco, J. (1991). *The Knowledge Link: How Firms Compete through Strategic Alliances*. Boston, MA: Harvard Business School Press.

Brown, J. (1989). *The Rational and the Social*. London: Routledge.

Byrne, B. M. (1988). Measuring Adolescent Self-Concept: Factorial Validity and Equivalency of the SDQ III Across Gender. *Multivariate Behavioral Research*, *23*(7), 361–375. doi:10.1207/s15327906mbr2303_5

Campbell, D. T., & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, *56*(2), 81–105. doi:10.1037/h0046016

Chong, C. W., Holden, T., Wilhelmij, P., & Schmidt, R. A. (2000). Where Does Knowledge Management Add Value? *Journal of Intellectual Capital*, *1*(4), 366–380. doi:10.1108/14691930010359261

Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *JMR, Journal of Marketing Research*, *16*(1), 64–73. doi:10.2307/3150876

Cohen, A. R., & Bradford, D. L. (1989). Influence without Authority: The Use of Alliances, Reciprocity and Exchange to Accomplish Work. *Organizational Dynamics*, *17*(3), 4–18. doi:10.1016/0090-2616(89)90033-8

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Company.

Corlett, J. A. (1996). *Analyzing Social Knowledge*. Totowa: Rowman & Littlefield Publishers.

Corlett, J. A. (2007). Analyzing Social Knowledge. *Social Epistemology*, *21*(3), 231–247. doi:10.1080/02691720701674049

Davenport, T. H., & Prusak, L. (2000). *Working Knowledge: How Organizations Manage what they Know*. Cambridge, MA: Harvard Business School Press.

Davenport, T.H., & Short, J.E. (1990). The New Industrial Engineering: Information Technology and Business Process Redesign. *Sloan Management Review, 31*(summer), 11-27.

Davidow, W. H., & Malone, M. (1992). *The Virtual Corporation*. London: Harper Collins.

Despres, C., & Chauvel, D. (2000). A Thematic Analysis of the Thinking in Knowledge Management. In Despres, C., & Chauvel, D. (Eds.), *Knowledge Horizons: The Present and the Promise of Knowledge Management*. Boston, MA: Butterworth-Heinemann.

Draper, N. R., & Smith, H. (1980). *Applied Regression Analysis* (2nd ed.). New York: John Wiley & Sons.

Firestone, J. M. (2001). Estimating Benefits of Knowledge Management Initiatives: Concepts, Methodology and Tools. *Journal of the KMCI*, *1*(3), 110–129.

Fukuyama, F. (1999). *The Great Disruption*. Simon & Schuster.

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. *Journal of Management Information Systems*, *18*(1), 185–214.

Goldman, A. (1995). Psychological, Social and Epistemic Factors in the Theory of Science. In M. Forbes & R. Burian (Eds.), *PSA 1994: Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association* (pp. 277-286). East Lansing, MI: Philosophy of Science Association.

Grant, R. M. (1996a). Prospering in Dynamically-competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science*, *7*(4), 375–387. doi:10.1287/orsc.7.4.375

Grant, R. M. (1996b). Towards a Knowledge-based Theory of the Firm. [Special Issue entitled Knowledge and the Firm]. *Strategic Management Journal*, *17*(4), 109–122.

Grant, R. M. (1997). Knowledge-based View of the Firm: Implications for Management Practice. *Long Range Planning*, *30*(3), 450–454. doi:10.1016/S0024-6301(97)00025-3

Grant, R. M., & Baden-Fuller, C. (1995). A Knowledge-based Theory of Inter-firm Collaboration. In *Academy of Management Best Papers Proceedings.*

Hansen, M. T., & von Oetinger, B. (2001). Introducing T-Shaped Managers*:* Knowledge Management's Next Generation. *Harvard Business Review*, *79*(3), 107–116.

Jovchelovitch, S. (2007). *Knowledge in Context: Representations, Community and Culture*. London: Routledge.

Kingsley, M. (2002). Measuring the Return on Knowledge Management. Retrieved on July 14, 2004, from http//:www.llrx.com/features/kmroi.html

Knight, D. J. (1999). Performance Measures for Increasing Intellectual Capital. *Strategy and Leadership*, *27*(2), 22–27. doi:10.1108/eb054632

Larsen, H. T., Bukh, P. N. D., & Mouritsen, J. (1999). Intellectual Capital Statements and Knowledge Management: 'Measuring' 'Reporting' 'Acting'. *Australian Accounting Review*, *9*(3), 15–26. doi:10.1111/j.1835-2561.1999.tb00113.x

Laudan, L. (1984). The Pseudo-Science of Science? In Brown, J. (Ed.), *Scientific Rationality: The Sociological Turn* (pp. 41–74). Dordrecht, Holland: Reidel.

Lee, H., & Choi, B. (2003). Knowledge Management Enablers Processes and Organizational Performance: An Integrative View and Empirical Study. *Journal of Management Information Systems*, *20*(1), 179–228.

Liebowitz, J., Rubenstein-Montano, B., McCaw, D., Buchwalter, J., Browning, C., Newman, B., & Rebeck, K. Knowledge Management Methodology Team. (2000). The Knowledge Audit. *Knowledge and Process Management*, (1): 3–10. doi:10.1002/(SICI)1099-1441(200001/03)7:1<3::AID-KPM72>3.0.CO;2-0

Madnick, S. E. (1991). The Information Technology Platform. In Morton, S. (Ed.), *The Corporation of the1990s* (pp. 27–60). New York: Oxford University Press.

Nelson, K. M., & Cooprider, J. G. (1996). The Contribution of Shared Knowledge to IS Group Performance. *Management Information Systems Quarterly*, *20*(4), 409–429. doi:10.2307/249562

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models*. USA: Irwin.

Nonaka, I. (1991). The Knowledge-Creating Company. *Harvard Business Review*, *69*(6), 96–104.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company*. Boston, MA: Oxford University Press.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Pan, S., & Scarbrough, H. (1998). A Socio-technical View of Knowledge-sharing at Buckman Laboratories. *Journal of Knowledge Management*, *2*(1), 55–66. doi:10.1108/EUM0000000004607

Papoutsakis, H., & Salvador Valles, R. (2006). Linking Knowledge Management and Information Technology to Business Performance: A Literature Review and a Proposed Model. *Journal of Knowledge Management Practice*, *7*(1).

Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research*. New York: CBS College Publishing.

Phillips, L. W., & Bagozzi, R. P. (1986). On Measuring Organizational Properties of Distributional Channels: Methodology Issues in the Use of Key Informants. *Research in Marketing*, *8*, 313–369.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.

Riggins, F. G., & Rhee, H. (1999). Developing the Learning Network Using Extranets. *International Journal of Electronic Commerce*, *4*(1), 65–83.

Roberts, J. (2000). From Know-how to Show-how?: Questioning the Role of Information and Communication Technologies in Knowledge Transfer. *Technology Analysis and Strategic Management*, *12*(4), 429–443. doi:10.1080/713698499

Sveiby, K. E. (1997). *The New Organizational Wealth: Managing and Measuring Knowledge-Based Assets*. Berrett-Koehler Publishers Inc.

Sveiby, K. E. (2001). A Knowledge-based Theory of the Firm To Guide Strategy Formulation. *Journal of Intellectual Capital*, *2*(4), 344–358. doi:10.1108/14691930110409651

Szulanski, G. (1996). Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within the Firm. *Strategic Management Journal*, *17*(10), 27–43.

Turiel, E. (1983). *The Development of Social Knowledge: Morality & Convention*. Cambridge, UK: Cambridge University Press.

Venkatraman, N. (1989). The Concept of Fit in Strategy Research: Toward Verbal and Statistical Correspondence. *Academy of Management Review*, *14*(3), 423–444. doi:10.2307/258177

von Krogh, G., Ichigo, K., & Nonaka, I. (2000). *Enabling Knowledge Creation. How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation*. New York: Oxford University Press.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, *5*, 161–215. doi:10.1214/aoms/1177732676

Wright, S. (1971). Path Coefficients and Path Regressions: Alternative or Complementary Concepts? In Blalock, H. M. (Ed.), *Causal Models in the Social Sciences* (pp. 101–114). Chicago: Aldine Publishing Co.

Zucker, L. (1986). Production of Trust: Institutional Sources of Economic Structure 1840-1920. In B.M. Staw & L.L. Cummings, (Ed.), *Research in Organizational Behavior*, 8, 53-111.

KEY TERMS AND DEFINITIONS

Influence: Nelson and Cooprider (1996, p. 414) define mutual influence as "the ability of groups to affect the key policies and decisions of each other." As organizational groups engaged in joint work are often dependent upon each other, influence relationships are created. One way influence is developed, is through the law of reciprocity. People expect payback for contribution to an exchange. The perception of reciprocal benefits leads to mutual influence and success in future exchanges among the groups. In our study, trust and influence have been recognized as antecedents of shared knowledge.

Information Technology: Davenport & Short (1990, p. 11) define Information Technology (IT) as "…the capabilities offered by computers, software applications, and telecommunications" and further explain that "IT should be viewed as more than an automating or mechanizing force; it can fundamentally reshape the way business is done" (p. 12) and that "IT can make it possible for employees scattered around the world to work as a team" (p. 19). Applegate, McFarlan & McKenney (1999; p. vii) identify IT as: "…computing, communications, business solutions and services…" and further down (note in p. 3) they explain that "…IT refers to technologies of computers and telecommunications (including data, voice, graphics, and full motion video)."

Performance: Under an industrial business management approach, manufacturing performance has three main activities: (i) the selection of goals; (ii) the consolidation of measurement information relevant to an organization's progress against these goals, and (iii) the interventions made by managers in light of this information with a view to improving future performance against these goals. Although presented here sequentially, typically all three activities will run concurrently, with the interventions made by managers affecting the choice of goals, the measurement information monitored, and the activities being undertaken within the organization.

Scientific Knowledge: Science is the process used everyday to logically complete thoughts through inference of facts determined by calculated experiments. As science itself has developed, the so produced scientific knowledge has developed a broader usage within scientists. The development of scientific methods has made a significant contribution to our understanding of scientific knowledge. To be termed scientific, a method of inquiry must be based on the collection of data through observation and experimentation, and the formulation and testing of hypotheses.

Shared Knowledge: "… an understanding and appreciation among [collaborating] groups and their managers, for the technologies and processes that affect their mutual performance" (Nelson & Cooprider 1996, p. 411). Appreciation and understanding are the two core elements of shared knowledge. Appreciation among diverse groups must be characterized by sensitivity to the point of view and interpretation of the other group, in order to overcome the barriers caused by the different environments and languages used. A deeper level of knowledge must be shared in order to achieve mutual understanding and this is often characterized as organizational knowledge Badaracco (1991).

Social Knowledge: Individuals develop social knowledge through their interactions with the social environment. Stable systems of social knowledge are organized around certain domains; the collaborating groups in our study. According to Turiel (1983) the acquisition of social knowledge can be interpreted in two different ways: (i) it can be knowledge transmitted to the individual by other persons, and in this case the knowledge acquired is dependent on what is transmitted; or (ii) it can be knowledge constructed by individuals specifically about certain social phenomena. The social dimension of scientific knowledge is of significant importance, as well. We perceive the social character of science as a matter of the aggregation of individuals, not their interactions, and social knowledge as simply the additive outcome of mostly scientists, members of the three groups, making sound scientific judgments.

Trust: has been defined as "a set of expectations shared by all those in an exchange" (Zucker, 1986) or as "the expectation shared by the [involved] groups that they will meet their commitments to each other" (Nelson and Cooprider, 1996, p. 413) or finally as "… maintaining reciprocal faith in each other in terms of intention and behaviors" (Lee and Choi, 2003, p. 190). The significance of trust has been given considerable attention and has even been described as a 'business imperative' (Davidow and Malone, 1992; Drucker, 1993 among others).

APPENDIX A: QUESTIONNAIRES AND CONSTRUCT MEASUREMENT

For reasons of economy of space the three Questionnaires (Relationship Questionnaires Type A and B and Performance Questionnaire Type C) are not presented separately. All the questions are listed, with their indicative number, upon analyzing the Indicators used for each Construct Measurement. In every question below, titles in brackets were customized to reflect the exact names of the participating organizations and functional groups, as they are used in every firm.

1. **Relationship Questionnaires** (Type A and B) included twelve questions aiming to measure:
   - Dependent or mediating variable Sharing Knowledge (3 questions)
   - Independent variable Mutual Trust (2 questions)
   - Independent variable Mutual Influence (4 questions)
   - The role and level of contribution of Information Technology (ITsk), both as a tool and/or enabler in supporting sharing knowledge among Manufacturing, Quality and/or R&D groups (2 questions)
   - The use of IT infrastructure –under the above described concept (1 question with multiple sub questions. Results are given in pie-chart form and are not presented here.)

Please characterize the general working relationship that currently exists between the

[Manufacturing] group and the [Quality or R&D] group → (Questionnaire Type A), or
[Quality or R&D] group and the [Manufacturing] group → (Questionnaire Type B).

Use Table 2 to measure constructs.

Table 2.

1	2	3	4	5	6	7
Extremely Weak	Weak	Moderately Weak	About Average	Moderately Strong	Strong	Extremely Strong

Shared Knowledge*

The three indicators of shared knowledge have been designed to assess the level of understanding or appreciation which the members of the three groups have of each others' work environments. Indicators 1 and 3 assess the level of appreciation that each participant has for what their partners (in the other group) have accomplished, by using general and multiplicative assessments respectively. The second indicator measures the level of understanding that the members of the three groups have of each others' work environments.

Shared Knowledge Indicator 1: (General Assessment, **Mean 5.2991**; SD 0.6957; Range 4)

A1/B1: The level of appreciation that the [Manufacturing] group and the [Quality or R&D] group have for each other's accomplishments is:

A1: (**Mean 5.35714**; SD 0.79250; Range 4)

B1: (**Mean 5.24107**; SD 0.84091; Range 4)

Shared Knowledge Indicator 2: (Multiplicative Assessment, **Mean 25.152**; SD 8.604; Range 44)

The product of the responses to the following:

A2: The level of understanding of the [Quality or R&D] group for the work environment (problems, tasks, roles, etc) of the [Manufacturing] group is:

(**Mean 4.84821**; SD 1.10045; Range 6)

B2: The level of understanding of the [Manufacturing] group for the work environment (problems, tasks, roles, etc) of the [Quality or R&D] group is:

(**Mean 5.17857**; SD 0.91252; Range 5)

Shared Knowledge Indicator 3: (Multiplicative Assessment, **Mean 26.652**; SD 8.157; Range 40)

The product of the responses to the following:

A3: The level of appreciation that the [Quality or R&D] group has for the accomplishments of the [Manufacturing] group is:

(**Mean 5.07143**; SD 0.97458; Range 4)

B3: The level of appreciation that the [Manufacturing] group has for the accomplishments of the [Quality or R&D] group is:

(**Mean 5.17857**; SD 0.91252; Range 5)

Shared Knowledge Construct: The mean of the above indicators (**Mean 19.034**; SD 5.180; Range 23.667).

Mutual Trust*

The two indicators of predisposition measure the extent to which the two partner groups trust each other. The first indicator directly assesses the level of trust between the groups, through a general assessment. The second indicator is a multiplicative assessment that evaluates the reputation of each group for meeting its commitments.

Mutual Trust Indicator 1: (General Assessment, **Mean 5.4509**; SD 0.8620; Range 4)

A4/B4. The level of trust that exists between the [Manufacturing] group and the [Quality or R&D] group is:

A4: Mean 5.54464; SD 1.10599; Range 5

B4: Mean 5.35714; SD 0.92860; Range 4

Mutual Trust Indicator 2: (Multiplicative Assessment, **Mean 28.304**; SD 8.374; Range 43)

The product of the responses to the following:

A5: The reputation of the [Quality or R&D] group for meeting its commitments to the [Manufacturing] group is: **Mean 5.44643**; SD 0.96646; Range 4
B5: The reputation of the [Manufacturing] group for meeting its commitments to the [Quality or R&D] group is: **Mean 5.13393**; SD 0.97256; Range 6
Mutual Trust Construct: The mean of the above indicators, **Mean 16.877**; SD 4.452; Range 21.5.

Mutual Influence*

The three indicators of mutual influence assess the level of influence and the ability to affect that members of the groups have on each others' key decisions and policies. The first indicator directly assesses the level of influence and the ability to affect between the groups, through a general assessment. The second indicator is a multiplicative assessment that evaluates the level of influence that the members of the groups have on each other's key decisions and policies. The third indicator is a multiplicative assessment that evaluates the ability to affect that the members of the groups have on each other's key decisions and policies

Mutual Influence Indicator 1: (General Assessment, **Mean 4.8973**; SD 0.7478; Range3.75)

The average of the responses to the following:

A6/B6: In general, the level of influence that members of the [Manufacturing] Group and the [Quality or R&D] have on each other's key decisions and policies is:
A6: Mean 5.01786; SD 0.97705; Range 5
B6: Mean 4.85714; SD 0.98509; Range 5
A7/B7: In general, the ability of members of the [Manufacturing] group and the [Quality or R&D] group to affect each other's key decisions and policies is:
A7: Mean 5.00000; SD 1.04838; Range 5
B7: Mean 4.71429; SD 1.06904; Range 5
Mutual Influence Indicator 2: (Multiplicative Assessment, **Mean 22.089**; SD 7.986; Range 33)

The product of the responses to the following:

A8: In general, the level of influence that members of the [Quality or R&D] group have on key decisions and policies of the [Manufacturing] group is:
Mean 4.81250; SD 0.92543; Range 4
B8: In general, the level of influence that members of the [Manufacturing] group have on key decisions and policies of the [Quality or R&D] group is:
Mean 4.50893; SD 1.17017; Range 6
Mutual Influence Indicator 3: (Multiplicative Assessment, **Mean 22.911**; SD 7.905; Range 33)

The product of the responses to the following:

A9: In general, the ability of members of the [Quality or R&D] group to affect key policies and decisions of the [Manufacturing] group is:
Mean 4.93750; SD 0.84129; Range 3
B9: In general, the ability of members of the [Manufacturing] group to affect key policies and decisions of the [Quality or R&D] group is:
Mean 5.57143; SD 1.19845; Range 5
Mutual Influence Construct: The mean of the above indicators, **Mean 16.632**; SD 5.099; Range 22.750.

(*) Questionnaire items for shared knowledge, mutual trust and mutual influence used in our study had been validated and used by Nelson and Cooprider (1996) upon exploring the concept of shared knowledge between Information Systems (IS) groups and their line customers as a contributor to IS performance.

Information Technology and Sharing Knowledge (ITsk)

By means of the relationship questionnaires (Type A and B) we are measuring the role and level of contribution of IT in supporting shared knowledge. We, thus, use the marker (sk) to distinguish from the IT indicators used in the performance questionnaire.

ITsk Indicator 1: (Multiplicative Assessment, **Mean 27.732**; SD 8.514; Range 40)

The product of the responses to the following:

A.10: In general, the role and the level of contribution of Information Technology (IT) as a tool and/or enabler, in supporting shared knowledge between [Manufacturing] group and [Quality or R&D] group is: (**Mean 5.25893**; SD 0.8776; Range 4)
B.10: In general, the role and the level of contribution of Information Technology (IT) as a tool and/or enabler, in supporting shared knowledge between [Quality or R&D] group and [Manufacturing] group is: (**Mean 5.19820**; SD 1.10223; Range 5)
ITsk Indicator 2: (Multiplicative Assessment, **Mean 29.223**; SD 8.379; Range 33)

The product of the responses to the following:

A.11: In general, the use of the Information Technology (IT) infrastructure in the [Manufacturing] group is: (**Mean 5.21429**; SD 0.90473; Range 5)
B.11: In general, the use of the Information Technology (IT) infrastructure in the [Quality or R&D] group is: (**Mean 5.54128**; SD 0.95774; Range 4)
Information Technology and Sharing Knowledge Construct (ITskC): The mean of the above indicators, Mean 28.478; SD 7.601; Range 34.

2. **Performance Questionnaire** (Type C) included nine questions aiming to measure:
   - Operational manufacturing performance (3 questions)
   - Service manufacturing performance (3 questions)

- The level of contribution of Information Technology (ITmp) to Manufacturing group performance (2 questions)
- The use of IT functions –under the above described concept (1 question with multiple sub questions. Results are given in pie-chart form and are not presented here.)

The following questions ask you to compare the [Manufacturing] group to other such Manufacturing groups. In relation to other comparable groups you have observed, how does the [Manufacturing] group rate on the following:

Use Table 3 to measure constructs.

Table 3.

1	2	3	4	5	6	7
Non-Existent	Very Weak	Weak	About Average	Strong	Very Strong	Extremely Strong

Manufacturing Performance

The indicators used to measure the two constructs of manufacturing performance in our study, are given in detail, here below. For reasons related to our initial study, we treated the answers separately (A for Manufacturing and B for Quality or R&D stakeholders), although this does not affect results here. As in approximately 95 per cent of the manufacturing units under investigation, the two stakeholders that completed the performance questionnaire were related, one to Production and the second to Quality or R&D (in most cases Production or Quality Directors) we have used multiplicative assessments of interaction for the questions relating manufacturing performance to collaboration among the groups.

Operational Manufacturing Performance

Operational MP Indicator 1: (Multiplicative Assessment)

The product of the two stakeholders' responses (from Manufacturing and Quality or R&D) to the following:

C1: In general, the quality of the work produced by the [Manufacturing] group for the [Quality or R&D] group is:

CA1: Mean 5.29464; SD 0.77852; Range 4

CB1: Mean 5.50000; SD 0.69749; Range 3

Operational MP Indicator 2: (General Assessment)

The average of the responses to the following:

C2: In general, the ability of the [Manufacturing] group to meet its organizational commitments (such as project schedules and budget) is:

CA2: Mean 5.33929; SD 0.87563; Range 5
CB2: Mean 5.33929; SD 0.72972; Range 3
Operational MP Indicator 3: (General Assessment)

The average of the responses to the following:

C3: In general, the ability of the [Manufacturing] group to meet its goals is:
CA3: Mean 5.41964; SD 0.74300, Range 3
CB3: Mean 5.37500; SD 0.77256; Range 3
Operational MP Construct: The mean of the above indicators, **Mean 13.385**; SD 2.641; Range 14.333.

Service Manufacturing Performance

Service MP Indicator 1: (Multiplicative Assessment)

The product of the two stakeholders' responses (from Manufacturing and Quality or R&D) to the following:

C4: In general, the ability of the [Manufacturing] group to react quickly to the [Quality or R&D] group's changing business needs is:
CA4: Mean 5.29464; SD 0.92647; Range 4
CB4: Mean 5.41964; SD 0.71834; Range 4
Service MP Indicator 2: (Multiplicative Assessment)

The product of the two stakeholders' responses (from Manufacturing and Quality or R&D) to the following:

C5: In general, the responsiveness of the [Manufacturing] group to the [Quality or R&D] group is:
CA5: Mean 5.18750; SD 0.92543; Range 4
CB5: Mean 5.27027; SD 0.79711; Range 4
Service MP Indicator 3: (Multiplicative Assessment)

The product of the two stakeholders' responses (from Manufacturing and Quality or R&D) to the following:

C6: In general, the contribution that the [Manufacturing] group has made to the accomplishment of the [Quality or R&D] group's strategic goals is:
CA6: Mean 5.41071; SD 0.95441; Range 5
CB6: Mean 5.25893; SD 0.86728; Range 4
Service MP Construct: The mean of the above indicators, **Mean 28.591**; SD 7.294; Range 37.667.
Manufacturing Performance Construct: The mean of Operational MP and Service MP constructs, **Mean 20.988**; SD 4.658; Range 21.25.

Information Technology and Manufacturing Performance (ITmp)

By means of the performance questionnaire (Type C) we are measuring the role and level of contribution of IT in supporting the performance of the manufacturing group. We therefore use the marker (mp) to distinguish from the IT indicators used in the relationship questionnaires (Type A and B).

ITmp Indicator 1: (Multiplicative Assessment, **Mean 28.348**; SD 7.673; Range 41)
C.A7: In general, the level of the Information Technology (IT) Contribution to the [Manufacturing] group performance is: (**Mean 5.17857**; SD 0.91252; Range 5)
C.B7: In general, the level of the Information Technology (IT) Contribution to the [Manufacturing] group performance is: (**Mean 5.38393**; SD 0.72591; Range 4)
ITmp Indicator 2: (General Assessment, **Mean 5.3170**; SD 0.8383; Range 3.5)
CA/B8: In general, the use of the Information Technology (IT) infrastructure, among the three groups is: (**Mean 5.22321**; SD 0.94640; Range 4)
Information Technology and Manufacturing Performance Construct (ITmpC): The mean of the above indicators, Mean 16.833; SD 4.069; Range 21.75.

It is noticeable that no significant difference is observed between responders of questionnaires A and B, regarding questions C.1 to C.7. Questions CA/B.8, due to their nature, have been analyzed as one.

APPENDIX B: REGRESSIONS AND CONFIRMATORY TESTS

General Note: Symbols used in our study and in the MINITAB extracts, included in the Appendixes, correlate as following:

β = Coef, t = T, p = P, r = R-Sq, R^2 = R-Sq(adj), and **F = F**.

ANOVA Table symbols:

DF=Degrees of Freedom, SS=Sums of Squares, MS=Mean Squares (SSR = SS Residual, SSTO = SS Total)

First Regression: MPC vs (MTC, MIC, SKC, ITmpC)

The regression equation is

```
MPC=media(OMPC,SMPC) = 6.98 + 0.354 MTC=media(MT1,MT2)
                        - 0.0364 MIC=media(MI1,MI2,MI3)
                        + 0.225 SKC=media(SK1,SK2,SK3)
                        + 0.259 ITmpC=media(ITmp1,ITmp2)
Predictor                    Coef  SE Coef      T      P  VIF
Constant                    6.981    1.873   3.73  0.000
MTC=media(MT1,MT2)         0.3535   0.1136   3.11  0.002  2.1
```

```
MIC=media(MI1,MI2,MI3)     -0.03643  0.08470  -0.43  0.668  1.5
SKC=media(SK1,SK2,SK3)       0.2248   0.1034   2.17  0.032  2.3
ITmpC=media(ITmp1,ITmp2)    0.25948  0.09151   2.84  0.005  1.1
S = 3.72201   R-Sq = 38.5%   R-Sq(adj) = 36.2%
```

Analysis of Variance

```
Source           DF       SS      MS      F      P
Regression        4   926.38  231.60  16.72  0.000
Residual Error  107  1482.31   13.85
Total           111  2408.69
Source                    DF  Seq SS
MTC=media(MT1,MT2)         1  730.61
MIC=media(MI1,MI2,MI3)     1   12.91
SKC=media(SK1,SK2,SK3)     1   71.49
ITmpC=media(ITmp1,ITmp2)   1  111.38
```

Unusual Observations

```
Obs  MTC=media(MT1,MT2)  MPC=media(OMPC,SMPC)     Fit  SE Fit  Residual
 38                15.3                28.083  19.417   0.619     8.666
 58                 8.0                 9.250  17.651   1.010    -8.401
 59                18.0                13.583  22.143   0.849    -8.559
107                20.8                18.917  23.830   1.523    -4.913
Obs  St Resid
 38      2.36R
 58     -2.35R
 59     -2.36R
107     -1.45 X
```

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

Second Regression: SKC vs (MTC, MIC, ITskC)

The regression equation is

```
SKC=media(SK1,SK2,SK3) = 1.08 + 0.639 MTC=media(MT1,MT2)
                         + 0.258 MIC=media(MI1,MI2,MI3)
                         + 0.101 ITskC=media(ITsk1,ITsk2)
Predictor                    Coef  SE Coef     T      P  VIF
Constant                    1.078    1.594  0.68  0.500
MTC=media(MT1,MT2)        0.63865  0.08285  7.71  0.000  1.3
```

```
MIC=media(MI1,MI2,MI3)     0.25800  0.07177  3.59  0.000  1.3
ITskC=media(ITsk1,ITsk2)  0.10137  0.04486  2.26  0.026  1.1
S = 3.38672   R-Sq = 58.4%   R-Sq(adj) = 57.3%
```

Analysis of Variance

```
Source           DF       SS      MS      F      P
Regression        3  1739.43  579.81  50.55  0.000
Residual Error  108  1238.74   11.47
Total           111  2978.17
Source                     DF   Seq SS
MTC=media(MT1,MT2)          1  1496.66
MIC=media(MI1,MI2,MI3)      1   184.22
ITskC=media(ITsk1,ITsk2)    1    58.56
```

Unusual Observations

```
Obs  MTC=media(MT1,MT2)  SKC=media(SK1,SK2,SK3)     Fit  SE Fit  Residual
  3                17.5                   9.333  16.556   0.763    -7.222
 10                21.5                  30.333  23.139   1.005     7.195
 13                23.8                  17.833  25.144   0.603    -7.311
 42                15.0                  23.833  16.661   0.578     7.172
 48                15.5                  16.167  16.116   1.227     0.051
 58                 8.0                  14.167  11.786   1.121     2.380
 64                17.3                  10.000  20.044   0.461   -10.044
 68                10.5                   6.667  13.650   0.561    -6.984
 74                21.5                  15.333  24.367   0.585    -9.034
107                20.8                  25.833  18.535   1.201     7.299
Obs  St Resid
  3    -2.19R
 10     2.22R
 13    -2.19R
 42     2.15R
 48     0.02 X
 58     0.74 X
 64    -2.99R
 68    -2.09R
 74    -2.71R
107     2.30RX
```

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

Confirmatory Tests

1 Cronbach's *alphas*

Have been calculated, for all variables involved, according to the formula:

$$\alpha \equiv \frac{n}{n-1}\left[1 - \frac{\sum \sigma^2_{\chi_i}}{\sigma^2_x}\right]$$

Where for the variable: $\chi_1, \ldots, \chi_i, \ldots, \chi_n$
$\sigma^2_{\chi_i}$ = variance of χ_i and σ^2_x = variance of $x = \sum \chi_i$

Shared Knowledge (SKC) = 0.9980971

Mutual Trust (MTC) = 0.99893219
Mutual Influence (MTC) = 0.99789307
Information Technology (ITskC) = 0.78191053
Information Technology (ITmpC) = 0.99919877
Manufacturing Performance (MPC) = 0.99870396
Operational Manufacturing Performance (OMPC) = 0.99935936
Service Manufacturing Performance (SMPC) = 0.81379442
2 MTMM Correlation Matrix
Correlations: MT1; MT2; MI1; MI2; MI3; SK1; SK2; SK3; OMPC; SMPC; ITskC; ITmpC

	MT1	MT2	MI1	MI2
MT2=A5*B5	**0.682**			
MI1=media(MI	0.574	0.478		
MI2=A8*B8	0.260	0.327	**0.691**	
MI3=A9*B9	0.371	0.493	**0.737**	**0.714**
SK1=media(A1	0.581	0.612	0.583	0.400
SK2=A2*B2	0.608	0.569	0.485	0.375
SK3=A3*B3	0.612	0.650	0.603	0.373
OMPC=media(C	0.524	0.486	0.515	0.301
SMPC=media(C	0.457	0.506	0.477	0.163
ITskC=media(	0.279	0.287	0.338	0.156
ITmpC=media(	0.057	0.247	0.262	0.319

	MI3	SK1	SK2	SK3
SK1=media(A1	0.464			
SK2=A2*B2	0.449	**0.597**		
SK3=A3*B3	0.574	**0.767**	**0.603**	
OMPC=media(C	0.390	0.448	0.448	0.532
SMPC=media(C	0.303	0.395	0.351	0.490

ITskC=media(	0.335	0.348	0.273	0.407
ITmpC=media(	0.217	0.208	0.197	0.233
	OMPC	SMPC	ITskC	
SMPC=media(C	**0.691**			
ITskC=media(	0.390	0.281		
ITmpC=media(	0.471	0.284	**0.460**	

This work was previously published in Social Knowledge: Using Social Media to Know What You Know, edited by John P. Girard and JoAnn L. Girard, pp. 207-235, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 60
Group Decision Making for Advanced Manufacturing Technology Selection Using the Choquet Integral

Cengiz Kahraman
Istanbul Technical University, Turkey

Selçuk Çebi
Karadeniz Technical University, Turkey

İhsan Kaya
Selçuk University, Turkey

ABSTRACT

Advanced manufacturing technology (AMT) is defined as a modern method of production incorporating highly automated and sophisticated computerized design and operational systems. Hence, an investment decision to adopt AMT is a strategic decision. A group decision making process is stressful when group members have different views under multiple and conflicting criteria. Satisfying group members' opinions has a critical impact on a decision. In this chapter, a multiple criteria group decision making problem under a fuzzy environment is used for the selection among AMTs. Choquet integral methodology is used for this selection. A strategic investment problem of a company for a suitable Automated Storage/Retrieval System (AS/RS) is considered and discussed.

INTRODUCTION

The developments of science and technology have led to many new concepts and products, which are replacing the old ones. Flexibility, improvement in productivity and quality, faster response to market shifts, shorter throughput and lead time and savings in inventory and labor costs, enable customer demands to be met in a shorter time. Changing customer preferences and tastes oblige the manufacturer to change his products frequently. Increased consumer awareness has led to the manufacture of high-quality goods.

DOI: 10.4018/978-1-4666-1945-6.ch060

The manufacturing process has to be faster to meet market demands at the appropriate time and to overcome competition. All these factors have led to changes in manufacturing processes, which have prompted many manufacturers to adopt computer-integrated manufacturing, namely advanced manufacturing technology (Aravindan & Punniyamoorthy, 2002).

The term ‘‘advanced manufacturing technology’’ (AMT) is broadly defined to include any automated (usually computer oriented) technology used in design, manufacturing/service, and decision support. Components of AMT include computer-aided engineering, factory management and control systems, computer-integrated manufacturing processes, and information integration. Different researchers have defined AMT in various ways. Typical items constituting AMT include: robotics, automated guided vehicles (AGVs), computer numerically controlled machines (CNC) machines, flexible manufacturing systems (FMS), computer-aided design (CAD), computer-aided manufacturing (CAM), information technology, cellular manufacturing and the use of just-in-time (JIT)/kanban system in the plant (Das and Jayaram, 2003)

AMTs provide many important benefits such as greater manufacturing flexibility, reduced inventory, reduced floor space, faster response to shifts in market demand, lower lead times, and a longer useful life of equipment over successive generations of products. Like many real-world problems, the decision of investing in advanced manufacturing technology frequently involves multiple and conflicting objectives, such as minimizing costs, maximizing flexibility, minimizing machine down times and maximizing efficiency. Application of traditional capital budgeting methods does not fully account for the benefits arising from intangible factors of AMTs (Kahraman et al., 2000).

The rapid growth of the AMT industry is creating problems in new directions. Prospective firms face the situation of having to make a decision among several AMTs, all of which are capable of performing a specific task. Since the development and use of appropriate assessment approaches are crucial to ensuring that the analysis of each AMT project considers all benefits and costs, the selection of a suitable AMT is becoming a more and more complex problem. Therefore the evaluation of the AMT is a multiple criteria problem and group decision making is generally preferred to solve these kinds of problem (Chuu, 2009). Group decision making requires considering multiple perspectives obtained from a group consisting multiple members (Lu et al., 2007). A group decision is required in two situations: (1) when a problem becomes too complex such that the knowledge of a decision maker is inadequate, as in product design, investment decisions and supplier selection and (2) when there are conflicting ideas that influence the decision makers like presidential elections. While the first one is called cooperative group decision making, the second one is called non-cooperative decision making. The common features of both decision making problems is to satisfy multiple decision makers’ preferences. Therefore, there are several kinds of group decision making methods in the literature such as authority rule, majority rule, negative minority rule, ranking rule, and consensus rule.

In this chapter, a multiple criteria group decision making problem under fuzzy environment is discussed to select an appropriate AMT by using the Choquet integral since the problem is related to both subjective and objective criteria. For a long time, it has been recognized that an exact description of many real-life physical situations may be virtually impossible. This is due to the high degree of imprecision involved in real-world situations.

Zadeh (1965, 1968) in his seminal papers proposed fuzzy set theory as the means for quantifying the inherent fuzziness that is present in ill-posed problems. Fuzziness is a type of imprecision which may be associated with the sets in which there is no sharp transition from membership to nonmembership. Many problems in the real world deal with uncertain and imprecise data so conventional approaches cannot be effective to find the best solution. To cope with this uncertainty, fuzzy set theory has been developed as an effective mathematical algebra in a vague environment. Although humans are comparatively efficient in qualitative forecasting, they are not good at making quantitative predictions. Since fuzzy linguistic models permit the translation of verbal expressions into numerical ones, thereby dealing quantitatively with imprecision in the expression of the importance of each criterion, some methods based on fuzzy relations are used. When the system involves human subjectivity, fuzzy algebra provides a mathematical framework for integrating imprecision and vagueness into the models (Zimmermann, 1991; Kaya & Çınar, 2008). A major contribution of fuzzy set theory is its capability in representing vague knowledge. In multiple criteria decision making problems, data very often are imprecise and fuzzy. Hence, a decision maker may encounter difficulty in quantifying such linguistic statements so they can be used in deterministic decision making.

The rest of this chapter is organized as follows: In Section 2, the basics of group decision making and a literature review on AMTs are presented. The selection criteria are discussed in Section 3. The fundamentals of the Choquet integral are given in Section 4. A case study is presented in Section 5. Finally, concluding remarks are given in Section 6.

LITERATURE REVIEW: GROUP DECISION MAKING AND AMTS

Group decision making consists of multiple individuals interacting to reach a decision. Each decision maker (expert) may have unique motivations or goals and may approach the decision process from a different angle, but have a common interest in reaching eventual agreement on selecting the "best" option(s). To do this, experts have to express their preferences by means of a set of evaluations over a set of alternatives (Herrera-Viedma et al., 2007). Group decision-making studies reveal that differences in the process by which groups reached decisions may help explain the different findings, generally within the group decision making literature. In some cases, group members privately considered information about a problem and formed individual judgments before meeting as a group. In other cases, groups are exposed to problems for the first time as a group (Moon et al., 2003). The main characteristics of group decisions are described as follows (Lu et al., 2007): (1) A group performs a decision-making task. (2) A group decision covers the whole process of transfer from generating ideas for solving a problem to implementing solutions. (3) Group members may be located in the same or different places. (4) Group members may work at the same or different times. (5) Group members may work for the same or different departments or organizations. (6) The group can be at any managerial levels. (7) There can be conflicting opinions among group members in a group decision process. (8) The decision task might have to be accomplished in a short time. (9) Group members might not have complete information for decision tasks. (10) Some required data, information or knowledge for a decision may be located in many sources and some may be external to the organization. In the following, some studies on AMTs from the literature are briefly summarized. The results of the literature review will be summarized in a comparative manner later.

Demmel and Askin (1992) introduced a multiple objective decision model for the evaluation of AMT, whereby the AMT selection attributes were classified into three categories: (1) pecuniary objective, (2) strategic objective, and (3) tactical objective. The pecuniary objective is based upon discounted cash flow techniques. Shang and Sueyoshi (1995) proposed a selection procedure for an FMS employing the AHP, simulation, and data envelopment analysis (DEA). Sambasivarao and Deshmukh (1997) devised a decision support system for selection of AMT. The system provided the decision-makers with a base of knowledge and approaches to building a justification model, and emphasized that AMT attributes could be classified into two groups, namely, tangible (objective) and intangible (subjective). They suggested AHP, TOPSIS, and a linear additive utility model as alternative multi-attribute analysis methods. Small and Chen (1997) discussed the results of a survey conducted in the US that investigated the use of justification approaches for AMS. According to their findings, manufacturing firms using hybrid strategies, which employ both economic and strategic justification techniques, attain significantly higher levels of success from advanced technology projects.

Kotha and Swamidass (1998) hypothesized that nationality of the firm was an important factor in AMT use. To test the effect of the nationality variable on AMT use they compared the use of 18 AMTs in the US and Japan in an exploratory study using data from 160 US firms and 125 Japanese firms. There was clear evidence that the nationality of the firm was a factor in AMT use; that was AMT use is significantly different in the two countries. O'Kane et al. (2000) reported that simulation modeling could be used for the justification phase of AMT. Efstathiades et al. (2000) investigated the transfer and implementation processes of AMT in developing countries using the Cypriot manufacturing industry as a case study. They also addressed two critical steps: (1) the management processes followed during the transfer of technology into the manufacturing environment, and (2) the steps followed both before and after implementation and productive operation of the technologies. They indicated that despite the distance between the manufacturer and the technology suppliers, no difficulties were experienced in acquiring information about the available technologies and the suitability of these technologies for the specific manufacturing environment.

Arvanitis and Hollenstein (2001) empirically investigated the decisions of firms to adopt AMT based on a comprehensive specification of a rank model of technology adoption using firm-level data for Swiss manufacturing. Chan et al. (2001) presented an overview and guidance for manufacturing companies which planned to invest in AMT. They explained the reasons why the company might encounter problems while adopting AMT and explained the many suggestions offered by the relevant literature for improving the performance of evaluation in AMT investment. According to the four major steps in adopting AMT (i.e. strategic planning, justification, training and installation, and routinization and implementation), they aimed at assisting managers or investors to recognize problems at each step, thus offering appropriate ways to avoid and/or solve those problems. Yusuff et al. (2001) applied the Analytic Hierarchy Process (AHP) to determine the priority weights in predicting the success of AMT implementation because of its capability to structure complicated, multi-decision maker, multi-alternative and multi-attribute problems hierarchically. They applied AHP to forecast the success of AMT implementation by taking into account two possible outcomes, which were predicted according to seven influential factors and called the successful and unsuccessful implementation. Kengpol and O'Brien (2001) applied AHP and cost/benefit analysis to select an advanced technology by incorporating quantitative and qualitative factors. Karsak and Tolga (2001) proposed a fuzzy multiple attribute decision-making (FMADM) procedure for evaluating advanced manufacturing system investments. This approach applied fuzzy discounted cash flow analysis and decision-makers' linguistic assessments to the economic criterion and strategic criteria, respectively. The decision algorithm aggregated the experts' preference ratings for the economic and strategic criteria weights, and the suitability of AMS investment alternatives versus the selection criteria to calculate fuzzy suitability indices. The fuzzy indices were then used to rank the AMS investment alternatives. Abdel-Kader and Dugdale (2001) proposed a FMADM model for assessing AMT investments. This model applied the mathematics of the analytical hierarchy process and fuzzy set theory to aggregate the two major dimensions of financial and non-financial attributes. Gupta and Whitehouse (2001) examined the impact of size of an organization on the manufacturing strategy.

Aravindan and Punniyamoorthy (2002) developed a new model to justify the investment in AMT using the extended Brown–Gibson model which was developed for evaluating alternative plant locations using certain objective and subjective factors. They showed that investment in AMT was attractive if the benefits accruing from the subjective factors were considered. Yurdakul (2004) utilized a combination of AHP and goal programming for selection among computer-integrated manufacturing technologies. The model used AHP to determine attribute weights in goal programming. Sener and Karsak (2007) presented a new decision making approach based on fuzzy regression with non-symmetric coefficients and fuzzy optimization for FMS selection.

Sener and Karsak (2008) proposed a decision model based on fuzzy linear regression and fuzzy multiple objective programming for AMT selection. Thakur and Jain (2008) explored the issues of measurement and comparison of the current state of AMT adoption in India, including important information technology (IT) factors. Comparisons were made between Indian firms, firms in a developed country (Canada), and in a developing country (China) for a worldwide perspective.

Chang and Wang (2009) investigated the potential use of AHP by applying the consistent fuzzy preference relations to predict advanced AMT implementation. They pointed out that the AHP method was not efficient enough because it performed complex computation procedures in paired comparisons and obtaining consistency indicators. To reduce the judgment times and avoid checking inconsistency, they employed the consistent fuzzy preference relations, which inherited some advantages of AHP (distinct hierarchy, effective numerical assessment), to overcome certain drawbacks resulting from this conventional pairwise comparison approach. They also reviewed several methods which are utilized to fulfill the requirements of AMT selection, evaluation, assessment, justification and prediction issues, such as simple estimation methods, breakeven approaches, systems value analysis, scoring models, analytic hierarchy process, simulation and strategic methods. Chuu (2009) developed a fuzzy multiple attribute decision-making method for the group decision-making to improve advanced manufacturing technology selection process. In the proposed approach, a new fusion method of fuzzy information was developed for managing information assessed in different linguistic (multi-granularity linguistic term sets) and numerical scales. Also an application for the Taiwanese bicycle industry was employed. Costa and Lima (2009) presented the rationality for the organizational design development related to AMT adoption. For the development of the organizational design process, it was suggested a specified sequence of activities, in order to consider all the dimensions that define the content of the organizational design.

A result from the literature review is that the Choquet integral has not been used up to date although various MCDM methods have been used for the solution AMT selection problem. In this chapter, it is the first time that Choquet integral is used as a group decision making technique for the solution of the AMT selection problem.

In Table 1, the results of the literature review are summarized in a comparative manner with respect to methods, purpose, and criteria.

Table 1. Categorization and comparison of related works for AMT decision-making

The method or technique	Author(s)	Purpose	Criteria
Analytic hierarchy process (AHP)	Yurdakul (2004)	Selection	Intangible/tangible
	Punniyamoorthy and Ragavan (2003)	Selection	Intangible/tangible
	Yusuff et al. (2001)	Prediction	Intangible
	Kengpol and O'Brien (2001)	Selection	Intangible/tangible
	Mohanty and Deshmukh (1998)	Selection	Intangible/tangible
	Chan et al. (1999)	Selection	Intangible/tangible
	Albayrakoglu (1996)	Justification	
	Oeltjenbruns et al. (1995)	Selection	Intangible/tangible
	Weber (1993)	Selection	Intangible/tangible
	Datta et al. (1992)	Justification	Intangible/tangible
Fuzzy AHP	Chan et al. (2006)	Selection	Intangible/tangible
Fuzzy multiple attribute decision-making	Karsak and Tolga (2001)	Evaluation	Intangible/tangible
	Abdel-Kader and Dugdale (2001)	Selection	Intangible/tangible
Fuzzy Group Decision Making	Chuu (2009)	Selection	Intangible/tangible
	Chuu (2009a)	Selection	Intangible/tangible
Traditional financial justification method	Orr (2002)	Evaluation	Intangible/tangible
Decision support systems	Sambasivarao and Deshmukh (1997)	Selection	Intangible/tangible
Simulation Modeling	O'Kane et al. (2000)	Justification	
Discounted cash flow (DCF)	Aravindan and Punniyamoorthy (2002)	Justification	Tangible
	Demmel and Askin (1992)	Evaluation	Tangible
Data envelop analysis (DEA)	Talluri et al. (2000)	Selection	Tangible
	Talluri and Yoon (2002)	Selection	Tangible
	Shang and Sueyoshi (1995)	Selection	Tangible
Mixed nonlinear programming model	Verter and Dasci (2002)	Planning	
	Bokhorst et al. (2002)	Selection	Tangible
System wide benefits value analysis	Ordoobadi and Mulvaney (2001)	Selection	Tangible
Cluster analysis	Díaz et al. (2003)	Examining	Intangible/tangible
Integrated planning model	Chen and Small (1994)	Planning	Intangible
	Efstathiades et al. (2002)	Planning	Intangible/tangible
Real option valuation technique	MacDougall and Pike (2003)		Intangible/tangible
Fuzzy regression	Sener and Karsak (2007)	Selection	Intangible/tangible
	Sener and Karsak (2008)	Selection	Intangible/tangible

SELECTION CRITERIA FOR ADVANCED MANUFACTURING TECHNOLOGY

Many companies are currently strengthening their competitive positions by updating the technologies used in the manufacturing processes. The selection of the appropriate technology that achieves or matches with the organization objective must be made on the basis of sound decision-making process. The decision to implement AMT is a major decision for many organizations. The success or failure of the organization could be due to this decision, and therefore it is very important that proper consideration be given to all aspects of implementation before a final commitment is made. This is necessary to ensure that all the expected benefits of the decision are realized (Yusuff et al., 2001).

Chan et al. (1999) used the following criteria in Table 2, to evaluate the tangible and intangible benefits for the selection of AMT.

THE CHOQUET INTEGRAL

In this chapter, we are interested in using the Choquet integral which is a generalization of the Lebesgue integral, defined with respect to a non classical measure, often called fuzzy measure, or non-additive measure or also capacity. When the underlying universe is finite, the Lebesgue integral reduces to a (convex) linear combination hence can be assimilated to a particular class of regression models, where the coeffcients are all positive and sum up to one. Hence, the Choquet integral offers a more general model, more precisely, it offers a set of (convex) linear models, each of them being defined in a simplex (Grabisch et al. 2007). The Choquet integral was applied to multi-criteria evaluation by many researchers (Ishii & Sugeno, 1985; Onisawa et al., 1986; Labreuche & Grabisch, 2003; Kojadinovic, 2004; Karsak, 2005; Tsai & Lu, 2006; Grabisch et al., 2007; Berrah, et al., 2008a, 2008b; Schmitt et al., 2008; Saad et al., 2008; Narukawa & Murofushi, 2008; Shieh et al., 2009). The Choquet integral is a method which measures the expected utility of an uncertain event. In imprecise probability theory, the Choquet integral is used to calculate the lower expectation induced by a 2-monotone lower probability, or the upper expectation induced by a 2-alternating upper probability. A fuzzy integral is a sort of general averaging operator

Table 2. Criteria to evaluate AMT Investments

Criteria	Sub criteria
1. Cost	• Product cost
	• Maintenance cost
	• High rate of return
	• Labor cost
	• Material cost
2. Performance	• Compatibility with existing machine
	• Work morale
	• Productivity
	• Utilization
	• Machine breakdown
	• Human integration
3. Quality	• Scraped value
	• Rework
	• Conformance
	• Consistency
4. Delivery	• Transportation
	• Customer services
	• Time scheduling
	• Delivery time
	• Inventory/work in progress
	• Lead time
5. Flexibility	• Design change accommodation
	• Change in product mix
	• Market responsiveness
	• Capacity growth
	• Routing and scheduling flexibility
6. Innovativeness	• Research and development
	• Introduce product variation

that can represent the notions of importance of a criterion and interactions among criteria. The most important feature of a fuzzy integral is its ability to represent a certain kind of interaction among criteria, ranging from redundancy (negative interaction) to synergy (positive interaction). According to Grabisch (1998), there is almost no well established method to deal with interacting criteria, and usually people tend to avoid the problem by constructing independent criteria (or criteria that are supposed to be so). This ability of the fuzzy integral is the reason for its success in various fields of multi-criteria decision-making. The disadvantage of fuzzy integral is complexity of the model, since the number of coefficients involved in a fuzzy integral model grows exponentially with the number of criteria to be aggregated. The main difficulty is to identify all these coefficients, either by some learning data, or by a questionnaire, or both.

To define fuzzy integrals, a set of values of importance is needed. This set is composed of the values of a fuzzy measure. So, a value of importance for each subset of attributes is needed. In the following, some definitions are given to explain the basics of the Choquet integral (Modave & Grabish, 1998):

Definition 1. Let I be the set of attributes (or any set in a general setting). A set function μ: $P(I) \rightarrow [0, 1]$ is called a fuzzy measure if it satisfies the three following axioms:

- $\mu(\emptyset) = 0$: an empty set has no importance,
- $\mu(I) = 1$: the maximal set has a maximal importance,
- $\mu(B) \leq \mu(C)$ if $B, C \subset I$ and $B \subset C$: a new added criterion cannot make the importance of a coalition (a set of criteria) diminish.

Therefore, in a problem where $card(I) = n$, a value for every element of $P(I)$ including 2^n values is needed. Assuming that the values of the empty set and of the maximal set are fixed, $(2^n\text{-}2)$ values or coefficients to define a fuzzy measure are needed. So, there is clearly a trade-off between complexity and accuracy. However, the complexity can be significantly reduced in order to guarantee that fuzzy measures are used in practical applications. A fuzzy integral is a sort of weighted mean taking into account the importance of every coalition of criteria.

Definition 2. Let μ be a fuzzy measure on $(I,P(I))$ and an application $f: I \rightarrow \Re^+$. The Choquet integral of f with respect to μ is defined by:

$$(C)\int_I f d\mu = \sum_{i=1}^{n}\left(f\left(\sigma\left(i\right)\right) - f\left(\sigma\left(i-1\right)\right)\right)\mu\left(A_{(i)}\right) \tag{1}$$

where σ is a permutation of the indices in order to have

$$f\left(\sigma\left(1\right)\right) \leq \ldots \leq f\left(\sigma\left(n\right)\right),$$
$$A_{(i)} = \left\{\sigma\left(i\right),\ldots,\sigma\left(n\right)\right\} \quad \text{, by convention.}$$
$$and \quad f\left(\sigma\left(0\right)\right) = 0$$

It is easy to see that the Choquet integral is a Lebesgue integral up to a reordering of the indices. Actually, if the fuzzy measure μ is additive, then the Choquet integral reduces to a Lebesgue integral.

Now we will illustrate how fuzzy measures can be used in lieu of the weighted sum and other more traditional aggregation operators in a Multi-criteria decision making framework. It is shown in Modave and Grabisch (1998) that under rather general assumptions over the set of alternatives X, and over the weak orders $\succ_i$, there exists a unique fuzzy measure μ over I such that:

$$\forall x, y \in X,\; x \succeq y \Leftrightarrow u\left(x\right) \geq u\left(y\right) \tag{2}$$

where

$$u(x) = \sum_{i=1}^{n} \left[u_{(i)}\left(x_{(i)}\right) - u_{(i-1)}\left(x_{(i-1)}\right) \right] \mu\left(A_{(i)}\right) \quad (3)$$

which is simply the aggregation of the monodimensional utility functions using the Choquet integral with respect to μ.

The global importance of a criterion is given by evaluating what this criterion brings to every coalition it does not belong to, and averaging this input. This is given by the Shapley value or index of importance.

Definition 3. Let μ be a fuzzy measure over I. The Shapley value of index j is defined by:

$$v(j) = \sum_{B \subset I | (j)} \gamma_I(B) \left[\mu\left(B \cup \{j\}\right) - \mu(B) \right] \quad (4)$$

with $\gamma_I(B) = \frac{\left(|I| - |B| - 1\right)!.|B|!}{|I|!}$, $|B|$ denotes the cardinal of B.

The Shapley value can be extended to degree two, in order to define the indices of interactions between attributes.

Definition 4. Let μ be a fuzzy measure over I. The interaction index between i and j is defined by:

$$I(i,j) = \sum_{B \subset I | \{i,j\}} \xi_I(B) .\left(\mu\left(B \cup \{i,j\}\right) - \mu\left(B \cup \{i\}\right) - \mu\left(B \cup \{i\}\right) + \mu(B) \right) \quad (5)$$

with $\xi_I(B) = \frac{\left(|I| - |B| - 2\right)!.|B|!}{\left(|I| - 1\right)!}$.

The interaction indices belong to the interval [−1,+1] and:

- *I(i, j)* > 0 if the attributes *i* and *j* are complementary;
- *I(i, j)* < 0 if the attributes *i* and *j* are redundant;
- *I(i, j)* = 0 if the attributes *i* and *j* are independent.

Interactions of higher orders can also be defined, however we will restrict ourselves to second order interactions which offer a good trade-off between accuracy and complexity.

Definition 5. A fuzzy measure μ is called 2-additive if all its interaction indices of order equal or larger than 3 are null and at least one interaction index of degree two is not null.

In this particular case of 2-additive measures, the following theorem can be given:

Theorem. Let μ be a 2-additive measure. Then Choquet integral can be computed by:

$$(C)\int_I f d\mu = \sum_{I_{0J} \succ 0} \left(f(i) \wedge f(j)\right) I_{ij} + \sum_{I_{0J} \prec 0} \left(f(i) \vee f(j)\right) \left|I_{ij}\right| + \sum_{i=1}^{n} f(i) \left(I_i - \frac{1}{2} \sum_{j \neq i} \left|I_{ij}\right| \right) \quad (6)$$

Note that this expression justifies the above interpretation of interaction indices, as a positive interaction index corresponds to a conjunction (complementary) and a negative interaction index corresponds to a disjunction (redundant).

The success of a Choquet integral depends on an appropriate representation of fuzzy measures, which captures the importance of individual criteria or their combination. In this chapter, the generalized Choquet integral proposed by Auephanwiriyakul et al. (2002) will be used, in which measurable evidence is represented in terms of intervals, whereas fuzzy measures are real numbers, and is an extension of the standard Choquet integral. In contrast to Auephanwiriyakul et al. (2002), Tsai and Lu (2006) proposes another generalization that involves linguistic expressions as well as information fusion between criteria to overcome vagueness and imprecision of linguistic terms in questionnaires.

Steps of the Methodology

The methodology is composed of eight steps (Tsai and Lu, 2006):

Step 1: Given criterion i, respondents' linguistic preferences for the degree of importance, perceived performance levels of alternative locations, and tolerance zone are surveyed.

Step 2: In view of the compatibility between perceived performance levels and the tolerance zone, trapezoidal fuzzy numbers are used to quantify all linguistic terms in this study. Given respondent t and criterion i, linguistic terms for the degree of importance which shows the weight of a criterion is parameterized by $\tilde{A}_i^t = \left(a_{i1}^t, a_{i2}^t, a_{i3}^t, a_{i4}^t\right)$, perceived performance levels which indicates score of an alternative by $\tilde{p}_i^t = \left(p_{i1}^t, p_{i2}^t, p_{i3}^t, p_{i4}^t\right)$, and the tolerance zone which represents specifications of a criterion by $\tilde{e}_i^t = \left(e_{i1L}^t, e_{i2L}^t, e_{i3U}^t, e_{i4U}^t\right)$. In this case study, t=1,2,3,4,5, i=1,2,…,n_j, j=1,2,3,4, n_1=3, n_2=2, n_3=4, n_4=3; where n_j represents the number of criteria in dimension j.

Step 3: Average $\tilde{A}_i^t$, $\tilde{p}_i^t$ and $\tilde{e}_i^t$ into $\tilde{A}_i$, $\tilde{p}_i$, and $\tilde{e}_i$, respectively using Equation (7).

$$\tilde{A}_i = \frac{\sum_{t=1}^k \tilde{A}_i^t}{k} = \left(\frac{\sum_{t=1}^k a_{i1}^t}{k}, \frac{\sum_{t=1}^k a_{i2}^t}{k}, \frac{\sum_{t=1}^k a_{i3}^t}{k}, \frac{\sum_{t=1}^k a_{i4}^t}{k}\right) \tag{7}$$

Step 4: Normalize the location value of each criterion using Equation (8).

$$\tilde{f}_i = \underset{\alpha\in[0,1]}{\Big\|} \overline{f}_i^{\alpha} = \underset{\alpha\in[0,1]}{\Big\|} [f_{i,\alpha}^-, f_{i,\alpha}^+] \tag{8}$$

where $f_i \in F(S)$ is a fuzzy-valued function. $\tilde{F}(S)$ is the set of all fuzzy-valued functions $f, f_i^{\alpha} = \left[f_{i,\alpha}^-, f_{i,\alpha}^+\right] = \dfrac{\overline{p}_i^{\alpha} - \overline{e}_i^{\alpha} + [1,1]}{2}$, $\overline{p}_i^{\alpha}$ and $\overline{e}_i^{\alpha}$ are α-level cuts of $\tilde{p}_i$ and $\tilde{e}_i$ for all α=[0,1].

Step 5: Find the location value of dimension j using Equation (9).

$$(C)\int \tilde{f} d\tilde{g} = \underset{\alpha=[0,1]}{\Big\|} \left[(C)\int f_{\alpha}^- dg_{\alpha}^-, (C)\int f_{\alpha}^+ dg_{\alpha}^+,\right] \tag{9}$$

where $\overline{g}_i : P(S) \to I(R^+)$, $\overline{g}_i = [g_i^-, g_i^+]$, $\overline{g}_i^{\alpha} = [g_{i,\alpha}^-, g_{i,\alpha}^+]$, $\overline{f}_i : S \to I(R^+)$, and $f_i = [f_i^-, f_i^+]$ for i=1, 2, 3,..., $n_{j.}$

To be able to calculate this location value, a λ value and the fuzzy measures $g(A_{(i)})$, i=1,2,...,n, are needed. These are obtained from Equations (10-12) (Sugeno, 1974; Ishii and Sugeno, 1985):

$$g\left(A_{(n)}\right) = g\left(\left\{s_{(n)}\right\}\right) = g_n \tag{10}$$

$$g\left(A_{(i)}\right) = g_i + g\left(A_{(i+1)}\right) + \lambda g_i g\left(A_{(i+1)}\right), \quad \text{where} \quad 1 \le i < n \tag{11}$$

$$1 = g\left(S\right) = \begin{cases} \dfrac{1}{\lambda}\left\{\prod_{i=1}^{n}\left[1 + \lambda g\left(A_i\right)\right] - 1\right\} & \text{if } \lambda \neq 0 \\ \sum_{i=1}^{n} g\left(A_i\right) & \text{if } \lambda = 0 \end{cases} \tag{12}$$

where, $A_i \cap A_j = \varphi$ for all i, j = 1,2,3,…,n and $i \neq j$, and λ∈(-1,∞].

Step 6: Aggregate all dimensional performance levels of the location alternatives into overall performance levels, using a hierarchical process applying the two-stage aggregation process of the generalized Choquet integral.

Equation 13.

$$\left.\begin{array}{l} main\ \ criterion_{(1)} = (C)\int fdg \\ \quad\vdots \\ main\ \ criterion_{(m)} = (C)\int fdg \end{array}\right\rangle \quad V = (\mathrm{C})\int main\ \ criterion\ \ dg$$

This is represented in Equation (13). The overall performance levels yields a fuzzy number $\tilde{V}$.

Step 7: Assume that the membership of $\tilde{V}$ is $\mu_{\tilde{V}}(x)$; defuzzify the fuzzy number $\tilde{V}$ into a crisp value *v* using Equation (14) and make a comparison of the overall performance levels of alternative locations.

$$F\left(\tilde{A}\right) = \frac{a_1 + a_2 + a_3 + a_4}{4} \qquad (14)$$

Step 8: Compare weak and advantageous criteria among the locations alternatives using Equation (8).

A CASE STUDY

Flexible material flow, delivery to the point of use, small lots, short production lead times, and high inventory turnover characterize modern manufacturing and distribution operations. Therefore, an automated storage and retrieval system (AS/RS) is used to satisfy the necessities of modern manufacturing systems. AS/RS is a system which consists of a variety of computer-controlled methods for automatically placing and retrieving loads from specific storage locations. The high throughput and short processing times of Automated storage/retrieval systems, promote improved inventory management practices and rapid material handling, and play an essential role in integrated manufacturing and distribution systems. Additional AS/RS benefits generally include high space utilization, labour savings, protection from pilferage and damage, better safety, and real-time material control (Park, 2004). There are five main categories of AS/RS. These

Table 3. The criteria of AS/RS selection and their symbols

Criteria	Sub criteria
C_1. Cost	
	C_{11}. Product cost
	C_{12}. Maintenance cost
	C_{13}. Labor cost
	C_{14}. Material cost
	C_{15}. Operating cost
	C_{16}. Breakdown cost
C_2. Performance	
	C_{21}. Compatibility with existing storehouse
	C_{22}. Productivity
	C_{23}. Utilization
	C_{24}. Rate of breakdown
	C_{25}. Human integration
C_3. Quality	
	C_{31}.Service Time
	C_{32}. Consistency
C_4. Flexibility	
	C_{41}. Design change accommodation
	C_{42}. Change in product mix
	C_{43}. Market responsiveness
	C_{44}. Capacity
C_5. Innovativeness	
	C_{51}.Research and development
	C_{52}. Introduce product variation

Table 4. Individual importances of criteria, the tolerance intervals, and linguistic evaluations for alternatives

		Individual Importances of Criteria			Tolerance Intervals	Deep Line AS/ RS			Unit Load AS/ RS		
Criteria	**Sub criteria**	**E1**	**E2**	**E3**		**E1**	**E2**	**E3**	**E1**	**E2**	**E3**
C1. Cost		VI	I	VI							
	C11 Product cost	I	I	VI	[M, H]	H	VH	VH	M	H	M
	C12 Maintenance cost	I	M	I	[L, M]	H	VH	VH	M	M	H
	C14 Labor cost	M	I	M	[VL, H]	VL	VL	L	VL	L	L
	C15 Material cost	I	I	I	[M, H]	H	VH	H	M	H	M
	C16 Operating cost	I	VI	I	[L, H]	H	VH	VH	H	VH	H
	C17 Breakdown cost	VI	VI	VI	[L, H]	H	M	H	VH	H	VH
C2. Performance		M	I	M							
	C21. Compatibility with existing storehouse	I	I	VI	[M, VH]	H	VH	H	M	M	H
	C22. Productivity	I	VI	I	[M, VH]	VH	H	VH	H	M	H
	C23. Utilization	M	U	M	[L, H]	H	VH	VH	H	H	M
	C24. Rate of breakdown	M	M	M	[VL, L]	L	L	VL	VL	L	L
	C25. Human integration	M	I	U	[M, M]	M	M	H	VH	VH	H
C3. Quality		M	I	M							
	C31.Service Time	VI	I	I	[L, H]	VH	H	VH	VH	H	H
	C32. Consistency	I	I	I	[M, VH]	H	VH	VH	H	H	H
C4. Flexibility		M	M	I							
	C41. Design change accommodation	U	U	M	[M, H]	H	H	M	L	M	L
	C42. Change in product mix	VU	U	U	[M, H]	L	L	M	M	L	L
	C43. Market responsiveness	I	I	M	[M, H]	H	VH	H	M	M	H
	C44. Capacity	M	I	M	[L, H]	VH	H	VH	M	H	H
C5. Innovativeness		M	M	I							
	C51.Research and development	M	I	I	[L, H]	M	H	M	H	VH	H
	C52. Introduce product variation	M	M	M	[L, H]	L	L	M	M	M	L

Table 5. The relationship between trapezoidal fuzzy numbers and degrees of linguistic importances in a five-linguistic-term scale

The degrees of importance		Low/High Levels		Trapezoidal fuzzy numbers
Label	**Linguistic terms**	**Label**	**Linguistic Terms**	
VU	Very Unimportant	VL	Very Low	(0,0,0,0.3)
U	Unimportant	L	Low	(0.2,0.3,0.4,0.5)
M	Middle	M	Middle	(0.4,0.5,0.6,0.7)
HI	High Important	H	High	(0.6,0.7,0.8,0.9)
VI	Very Important	VH	Very High	(0.8,1,1,1)

are; *(i)* unit load AS/RS, *(ii)* mini load AS/RS, *(iii)* deep-lane AS/RS *(iv)* man-on-board AS/RS, and *(v)* automated item retrieval system. The unit load AS/RS is used to store and retrieve loads that are palletized or stored in standard-sized containers. The mini load AS/RS is designed to handle small loads such as individual parts, tools, and supplies that are contained in bins or drawers in the storage system. A mini load AS/RS is generally smaller than a unit load AS/RS and allows storing more material in less space. The deep-lane AS/RS is a high-density unit load storage system that is appropriate for storing large quantities of stock. The man-on-board AS/RS system allows storage of items in less than unit load quantities. The system permits individual items to be picked directly at their storage locations. The automated item retrieval system is designed for retrieval of individual items or small product cartoons.

In our case study, one of the biggest logistic firms in Turkey plans to install an AS/RS system into its warehouse. The firm will make a selection between a deep-lane AS/RS and a unit load AS/RS by using Choquet integral methodology.

Step 1: The criteria given in Table 3 are selected from the literature to evaluate the alternatives (Chan et al., 1999). Three experts assign their linguistic preferences as seen in Table 4.

Table 6. Aggregated decision matrix belonging to the group which consists of three experts

Criteria	Individual Importance	Tolerance Zone	Deep Line AS/RS	Unit Load AS/RS
C1	(0.733,0.9,0.933,0.967)			
C11	(0.667,0.8,0.867,0.933)	(0.4,0.5,0.8,0.9)	(0.733,0.9,0.933,0.967)	(0.467,0.567,0.667,0.767)
C12	(0.533,0.667,0.733,0.833)	(0.2,0.3,0.6,0.7)	(0.733,0.9,0.933,0.967)	(0.467,0.567,0.667,0.767)
C13	(0.467,0.633,0.667,0.767)	(0,0,0.8,0.9)	(0.067,0.1,0.133,0.367)	(0.133,0.2,0.267,0.433)
C14	(0.6,0.7,0.8,0.9)	(0.4,0.5,0.8,0.9)	(0.667,0.8,0.867,0.933)	(0.467,0.567,0.667,0.767)
C15	(0.667,0.8,0.867,0.933)	(0.2,0.3,0.6,0.7)	(0.733,0.9,0.933,0.967)	(0.667,0.8,0.867,0.933)
C16	(0.8,1,1,1)	(0.2,0.3,0.6,0.7)	(0.533,0.633,0.733,0.833)	(0.733,0.9,0.933,0.967)
C2	(0.467,0.567,0.667,0.767)			
C21	(0.667,0.8,0.867,0.933)	(0.4,0.5,1,1)	(0.667,0.8,0.867,0.933)	(0.467,0.633,0.667,0.767)
C22	(0.667,0.8,0.867,0.933)	(0.4,0.5,1,1)	(0.733,0.9,0.933,0.967)	(0.533,0.667,0.733,0.833)
C23	(0.267,0.4,0.4,0.567)	(0.2,0.3,0.8,0.9)	(0.733,0.9,0.933,0.967)	(0.533,0.667,0.733,0.833)
C24	(0.4,0.6,0.6,0.7)	(0,0,0.4,0.5)	(0.133,0.267,0.267,0.433)	(0.133,0.267,0.267,0.433)
C25	(0.333,0.433,0.467,0.633)	(0.4,0.5,0.6,0.7)	(0.467,0.633,0.667,0.767)	(0.733,0.9,0.933,0.967)
C3	(0.467,0.633,0.667,0.767)			
C31	(0.667,0.8,0.867,0.933)	(0.2,0.3,0.8,0.9)	(0.733,0.9,0.933,0.967)	(0.667,0.8,0.867,0.933)
C32	(0.6,0.7,0.8,0.9)	(0.4,0.5,1,1)	(0.733,0.9,0.933,0.967)	(0.6,0.7,0.8,0.9)
C4	(0.467,0.633,0.667,0.767)			
C41	(0.133,0.2,0.2,0.433)	(0.4,0.5,0.8,0.9)	(0.533,0.667,0.733,0.833)	(0.267,0.467,0.467,0.567)
C42	(0.067,0.133,0.133,0.367)	(0.4,0.5,0.8,0.9)	(0.267,0.467,0.467,0.567)	(0.267,0.467,0.467,0.567)
C43	(0.533,0.667,0.733,0.833)	(0.4,0.5,0.8,0.9)	(0.667,0.8,0.867,0.933)	(0.467,0.633,0.667,0.767)
C44	(0.467,0.633,0.667,0.767)	(0.2,0.3,0.8,0.9)	(0.733,0.9,0.933,0.967)	(0.533,0.667,0.733,0.833)
C5	(0.467,0.633,0.667,0.767)			
C51	(0.533,0.667,0.733,0.833)	(0.2,0.3,0.8,0.9)	(0.467,0.633,0.667,0.767)	(0.667,0.8,0.867,0.933)
C52	(0.4,0.6,0.6,0.7)	(0.2,0.3,0.8,0.9)	(0.267,0.467,0.467,0.567)	(0.333,0.533,0.533,0.633)

Step 2: Trapezoidal fuzzy numbers given in Table 5 are used to quantify the linguistic terms in Table 4.

Step 3: The aggregation of experts' assessments obtained by Equation (7) is given in Table 6.

Step 4: By using Equation (8), the evaluation values are normalized for each criterion.

Step 5: At different α levels, the values of all criteria in the same dimension are aggregated by using Equation (9). To illustrate the calculation procedure in Steps 4 and 5, an example is presented for deep line AS/RS alternative under innovativeness criterion. Using Equation (8),

$$f, f_i^{\alpha} = \left[f_{1,0}^{-}, f_{1,0}^{+}\right] = \frac{\left[0.667,\ 0.93\right] - \left[0.2,\ 0.9\right] + \left[1,\ 1\right]}{2} = \left[0.283,\ 0.783\right]$$

and $\left[f_{2,0}^{-}, f_{2,0}^{+}\right] = [0.183,\ 0.683]$ are obtained. Their corresponding degrees of importances are

Table 7. Defuzzified values of AS/RS alternatives

Criteria			Deep Line AS/RS	Unit Load AS/RS
	Deep Line AS/RS	Unit Load AS/RS	Rank1	Rank2
Overall Value	(0.468,0.631,0.811,0.882)	(0.458,0.634,0.812,0.882)	**0.698**	0.696
C1	(0.5,0.643,0.813,0.883)	(0.5,0.65,0.817,0.883)	0.71	**0.713**
C11	(0.417,0.55,0.717,0.783)	(0.284,0.383,0.583,0.683)	**0.617**	0.483
C12	(0.517,0.65,0.817,0.883)	(0.384,0.483,0.683,0.783)	**0.717**	0.583
C13	(0.084,0.15,0.567,0.683)	(0.117,0.2,0.633,0.717)	0.371	**0.417**
C14	(0.384,0.5,0.684,0.767)	(0.284,0.383,0.583,0.683)	0.583	**0.483**
C15	(0.517,0.65,0.817,0.883)	(0.484,0.6,0.783,0.867)	**0.717**	0.683
C16	(0.417,0.517,0.717,0.817)	(0.517,0.65,0.817,0.883)	0.617	**0.717**
C2	(0.379,0.506,0.753,0.839)	(0.366,0.513,0.687,0.792)	**0.619**	0.59
C21	(0.334,0.4,0.684,0.767)	(0.234,0.317,0.583,0.683)	**0.546**	0.454
C22	(0.367,0.45,0.717,0.783)	(0.267,0.333,0.617,0.717)	**0.579**	0.483
C23	(0.417,0.55,0.817,0.883)	(0.317,0.433,0.717,0.817)	**0.667**	0.571
C24	(0.317,0.434,0.634,0.717)	(0.317,0.433,0.633,0.717)	0.525	0.525
C25	(0.384,0.517,0.584,0.683)	(0.517,0.65,0.717,0.783)	0.542	**0.667**
C3	(0.397,0.52,0.797,0.873)	(0.35,0.455,0.757,0.855)	**0.647**	0.604
C31	(0.417,0.55,0.817,0.883)	(0.383,0.5,0.783,0.867)	**0.667**	0.633
C32	(0.367,0.45,0.717,0.783)	(0.3,0.35,0.65,0.75)	**0.579**	0.513
C4	(0.386,0.525,0.767,0.853)	(0.287,0.422,0.668,0.783)	**0.633**	0.54
C41	(0.317,0.434,0.617,0.717)	(0.184,0.333,0.483,0.583)	**0.521**	0.396
C42	(0.184,0.334,0.484,0.583)	(0.184,0.333,0.483,0.583)	0.396	0.396
C43	(0.384,0.5,0.684,0.767)	(0.284,0.417,0.583,0.683)	**0.583**	0.492
C44	(0.417,0.55,0.817,0.883)	(0.317,0.433,0.717,0.817)	**0.667**	0.571
C5	(0.233,0.383,0.643,0.753)	(0.283,0.447,0.717,0.822)	0.5	**0.567**
C51	(0.283,0.417,0.683,0.783)	(0.383,0.5,0.783,0.867)	0.542	**0.633**
C52	(0.183,0.333,0.583,0.683)	(0.217,0.367,0.617,0.717)	0.446	**0.479**

$\bar{g}_1^{\,0} = \left[0.533,\ 0.833\right],\ \bar{g}_2^{\,0} = \left[0.4,\ 0.6\right],$ respectively. First, the sequence $f_{i,0}^{\,-}$ is sorted, where i=1 and 2, as follows: $f_{2,0}^{-} < f_{1,0}^{-}$. The corresponding degrees of importances are $\bar{g}_{1,0} = 0.4$ and $\bar{g}_{2,0} = 0.533,$ respectively. By solving the equation $0 = \frac{1}{\lambda}\left\{\prod_{i=1}^{2}\left[1 + \lambda \bar{g}_i\right] - 1\right\},$ λ=0.313 is obtained. Then, their fuzzy measures are derived from as follow;

g(A(2)) = g2 = 0.533

g(A(1)) = g1 + g(A(2)) + λ g1 g(A(2)) = 1

The aggregated Choquet integral values for the main criterion C5 are calculated as $(C)\int \tilde{f} d\tilde{g} = \left[0.237,\ 0.753\right]$

Step 6: Similar to Steps 4 and 5, the overall values are obtained for AS/RS systems, as shown in Table 7.

Step 7: From Table 7, the defuzzified overall values of AS/RSs using generalized Chouqet Integral are obtained as 0.710 and 0.713. Hence Unit Load AS/RS is the best alternative for the firm.

Step 8: Weak and advantageous criteria for alternatives are clarified in Table 7 by normal and bold characters, respectively. The numbers in bold characters in Table 7 represent that the alternative has more advantage than the other for that criterion.

CONCLUSION

In this chapter, a multiple criteria group decision-making problem under a fuzzy environment has been discussed. To illustrate the chapter, a selection problem between AS/RSs has been taken into consideration by using the Choquet integral. The characteristic feature of the problem is that it includes both objective and subjective and conflicting criteria. Hence, three experts' expertise is used in order to cope with the complexity of the problem. To satisfy the group decision, the arithmetic mean method has been used. The Choquet integral has provided an excellent tool for the solution of our advanced manufacturing technology selection problem including conflicting criteria.

For future research, another aggregation method from the literature may be used in order to compare the results and put forward the consistency of the group decision.

REFERENCES

Abdel-Kader, M. G., & Dugdale, D. (2001). Evaluating investments in advanced manufacturing technology: A fuzzy set theory approach. *The British Accounting Review*, *33*, 455–489. doi:10.1006/bare.2001.0177

Albayrakoglu, M. (1996). Justification of new manufacturing technology: A strategic approach using the analytic hierarchy process. *Production and Inventory Management Journal*, *37*(1), 71–76.

Aravindan, P., & Punniyamoorthy, M. (2002). Justification of advanced manufacturing technologies (AMT). *International Journal of Advanced Manufacturing Technology*, *19*, 151–156. doi:10.1007/s001700200008

Arvanitis, S., & Hollenstein, H. (2001). The determinants of the adoption of advanced manufacturing technology. *Economics of Innovation and New Technology*, *10*(5), 377–414. doi:10.1080/10438590100000015

Auephanwiriyakul, S., Keller, J. M., & Gader, P. D. (2002). Generalized Choquet fuzzy integral fusion. *Information Fusion*, *3*, 69–85. doi:10.1016/S1566-2535(01)00054-9

Berrah, L., Mauris, G., Montmain, J., & Cliville, V. (2008a). Efficacy and efficiency indexes for a multi-criteria industrial performance synthesized by Choquet integral aggregation. *International Journal of Computer Integrated Manufacturing*, *21*(4), 415–425. doi:10.1080/09511920701574255

Berrah, L., Mauris, G., Montmain, J., & Cliville, V. (2008b). Monitoring the improvement of an overall industrial performance based on a Choquet integral aggregation. *Omega-International Journal of Management Science*, *36*(3), 340–351. doi:10.1016/j.omega.2006.02.009

Bokhorst, J. A. C., Slomp, J., & Suresh, N. C. (2002). An integrated model for part operation allocation and investments in CNC technology. *International Journal of Production Economics*, *75*(3), 267–285. doi:10.1016/S0925-5273(01)00118-9

Chan, F. T. S., Chan, H. K., Chan, M. H., & Humphreys, P. K. (2006). An integrated fuzzy approach for the selection of manufacturing technologies. *International Journal of Advanced Manufacturing Technology*, *27*(7–8), 747–758. doi:10.1007/s00170-004-2246-9

Chan, F. T. S., Chan, M. H., Lau, H., & Ip, R. W. L. (2001). Investment appraisal techniques for advanced manufacturing technology (AMT): A literature review. *Integrated Manufacturing Systems*, *12*(1), 35–47. doi:10.1108/09576060110361528

Chan, F. T. S., Chan, M. H., & Mak, K. L. (1999). An integrated approach to investment appraisal for advanced manufacturing technology. *Human Factors and Ergonomics in Manufacturing*, *9*(1), 69–86. doi:10.1002/(SICI)1520-6564(199924)9:1<69::AID-HFM4>3.0.CO;2-1

Chang, T. H., & Wang, T. C. (2009). Measuring the success possibility of implementing advanced manufacturing technology by utilizing the consistent fuzzy preference relations. *Expert Systems with Applications*, *36*, 4313–4320. doi:10.1016/j.eswa.2008.03.019

Chen, I. J., & Small, M. H. (1994). Implementing manufacturing technology: An integrated planning model. *Omega*, *22*(1), 91–103. doi:10.1016/0305-0483(94)90010-8

Chuu, S. J. (2009a). Selecting the advanced manufacturing technology using fuzzy multiple attributes group decision making with multiple fuzzy information. *Computers & Industrial Engineering*, *57*(3), 1033–1042. doi:10.1016/j.cie.2009.04.011

Chuu, S. J. (2009b). Group decision-making model using fuzzy multiple attributes analysis for the evaluation of advanced manufacturing technology. *Fuzzy Sets and Systems*, *160*, 586–602. doi:10.1016/j.fss.2008.07.015

Costa, S. E. G., & Lima, E. P. (2009). Advanced manufacturing technology adoption: An integrated approach. *Journal of Manufacturing Technology Management*, *20*(1), 74–96. doi:10.1108/17410380910925415

D’ıaz, M. S., Machuca, J. A. D., & Álvarez-Gil, M. J. (2003). A view of developing patterns of investment in AMT through empirical taxonomies: New evidence. *Journal of Operations Management*, *21*(5), 577–606. doi:10.1016/j.jom.2003.03.002

Das, A., & Jayaram, J. (2003). Relative importance of contingency variables for advanced manufacturing technology. *International Journal of Production Research*, *41*(18), 4429–4452. doi:10.1080/00207540310001595819

Datta, V., Sambasivarao, K. V., Kodali, R., & Deshmukh, S. G. (1992). Multi-attribute decision model using the analytic hierarchy process for justification of manufacturing systems. *International Journal of Production Economics*, *28*(2), 227–234. doi:10.1016/0925-5273(92)90035-6

Demmel, J. G., & Askin, R. G. (1992). A multiple-objective decision model for the evaluation of advanced manufacturing system technologies. *Journal of Manufacturing Systems*, *11*(3), 179–194. doi:10.1016/0278-6125(92)90004-Y

Efstathiades, A., Tassou, S., & Antoniou, A. (2002). Strategic planning, transfer and implementation of advanced manufacturing technologies (AMT): Development of an integrated process plan. *Technovation*, *22*, 201–212. doi:10.1016/S0166-4972(01)00024-4

Grabisch, M., Kojadinovic, I., & Meyer, P. (2007). A review of methods for capacity identification in Choquet integral based multi-attribute utility theory applications of the Kappalab R package. *European Journal of Operational Research*, *186*(2), 766–785. doi:10.1016/j.ejor.2007.02.025

Gupta, A., & Whitehouse, F. R. (2001). Firms using advanced manufacturing technology management: An empirical analysis based on size. *Integrated Manufacturing Systems*, *12*(5), 346–350. doi:10.1108/09576060110398500

Herrera-Viedma, E., Chiclana, F., Herrera, F., & Alonso, S. (2007). Group decision-making model with incomplete fuzzy preference relations based on additive consistency. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *37*(1), 176–189. doi:10.1109/TSMCB.2006.875872

Ishii, K., & Sugeno, M. (1985). A model of human evaluation process using fuzzy integral. *International Journal of Man-Machine Studies*, *22*(1), 19–38. doi:10.1016/S0020-7373(85)80075-4

Jahanshahloo, G. R., Lotfi, F. H., & Izadikhah, M. (2006). Extension of the TOPSIS method for decision-making problems with fuzzy data. *Applied Mathematics and Computation*, *181*, 1544–1551. doi:10.1016/j.amc.2006.02.057

Karsak, E., & Tolga, E. (2001). Fuzzy multi-criteria decision-making procedure for evaluating advanced manufacturing system investments. *International Journal of Production Economics*, *69*, 49–64. doi:10.1016/S0925-5273(00)00081-5

Karsak, E. E. (2005). Choquet integral-based decision making approach for robot selection. *Knowledge-Based Intelligent Information and Engineering Systems,* 635-641.

Kaya, İ., & Çınar, D. (2008). Facility location selection using a fuzzy outranking method. *Journal of Multiple-Valued Logic and Soft Computing*, *14*, 251–263.

Kengpol, A., & O'Brien, C. (2001). The development of a decision support tool for the selection of advanced technology to achieve rapid product development. *International Journal of Production Economics*, *69*, 177–191. doi:10.1016/S0925-5273(00)00016-5

Kojadinovic, I. (2004). Unsupervised aggregation by Choquet integral based on entropy functionals: Application to the evaluation of students. *Modeling Decisions for Artificial Intelligence*, *3131*, 163–174.

Kotha, S., & Swamidass, P. M. (1998). Advanced manufacturing technology use: Exploring the effect of the nationality variable. *International Journal of Production Research*, *36*(11), 3135–3146. doi:10.1080/002075498192337

Labreuche, C., & Grabisch, M. (2003). Choquet integral for the aggregation of interval scales in multicriteria decision making. *Fuzzy Sets and Systems*, *137*(1), 11–26. doi:10.1016/S0165-0114(02)00429-3

Lu, J., Zhang, G., Ruan, D., & Wu, F. (2007). *Multi-objective group decision making methods: Software and applications with fuzzy set techniques*. Singapore: Imperial Collage Press.

MacDougall, S. L., & Pike, R. H. (2003). Consider your options: Changes to strategic value during implementation of advanced manufacturing technology. *Omega*, *31*(1), 1–15. doi:10.1016/S0305-0483(02)00061-0

Modave, F., & Grabisch, M. (1998). Preference representation by Choquet integral: The commensurability hypothesis. *Proc. 7th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, 1-8.

Mohanty, R. P., & Deshmukh, S. G. (1998). Advanced manufacturing technology selection: A strategic model for learning and evaluation. *International Journal of Production Economics*, *55*(3), 295–307. doi:10.1016/S0925-5273(98)00075-9

Moon, H., Conlon, D. E., Humphrey, S. E., Quigley, N., Devers, C. E., & Nowakowski, J. M. (2003). Group decision process and incrementalism in organizational decision making. *Organizational Behavior and Human Decision Processes*, *92*(1-2), 67–79. doi:10.1016/S0749-5978(03)00079-7

Narukawa, Y., & Murofushi, T. (2008). Choquet-Stieltjes integral as a tool for decision modeling. *International Journal of Intelligent Systems*, *23*(2), 115–127. doi:10.1002/int.20260

O'Kane, J. F., Spenceley, J. R., & Taylor, R. (2000). Simulation as an essential tool for advanced manufacturing technology problems. *Journal of Materials Processing Technology*, *107*, 412–424. doi:10.1016/S0924-0136(00)00689-0

Oeltjenbruns, H., Kolarik, W. J., & Schnadt–Kirschner, R. (1995). Strategic planning in manufacturing systems: AHP application to an equipment replacement decision. *International Journal of Production Economics*, *38*, 189–197. doi:10.1016/0925-5273(94)00092-O

Onisawa, T., Sugeno, M., Nishiwaki, M. Y., Kawai, H., & Harima, Y. (1986). Fuzzy measure analysis of public attitude towards the use of nuclear energy. *Fuzzy Sets and Systems*, *20*, 259–289. doi:10.1016/S0165-0114(86)90040-0

Ordoobadi, S. M., & Mulvaney, N. J. (2001). Development of a justification tool for advanced manufacturing technologies: System-wide benefits value analysis. *Journal of Engineering and Technology Management*, *18*, 157–184. doi:10.1016/S0923-4748(01)00033-9

Orr, S. (2002). A comparison of AMT strategies in the USA, South Africa and Germany. *International Journal of Manufacturing Technology Management*, *4*(6), 441–454. doi:10.1504/IJMTM.2002.002517

Park, B. C. (2001). An optimal dwell point policy for automated storage/retrieval systems with uniformly distributed, rectangular racks. *International Journal of Production Research*, *39*(7), 1469–1480. doi:10.1080/00207540010023583

Punniyamoorthy, M., & Ragavan, P. V. (2003). A strategic decision model for the justification of technology selection. *International Journal of Advanced Manufacturing Technology*, *21*(1), 72–78. doi:10.1007/s001700300008

Saad, I., & Hammadi, S. (2008). Choquet integral for criteria aggregation in the flexible job-shop scheduling problems. *Mathematics and Computers in Simulation*, *76*(5-6), 447–462. doi:10.1016/j.matcom.2007.04.010

Sambasivarao, K. V., & Deshmukh, S. G. (1997). A decision support system for selection and justification of advanced manufacturing technologies. *Production Planning and Control*, *8*(3), 270–284. doi:10.1080/095372897235325

Schmitt, E., Bombardier, V., & Wendling, L. (2008). Improving fuzzy rule classifier by extracting suitable features from capacities with respect to Choquet integral. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *38*(5), 1195–1206. doi:10.1109/TSMCB.2008.925750

Sener, Z., & Karsak, E. E. (2007). A decision model for advanced manufacturing technology selection using fuzzy regression and fuzzy optimization. *IEEE International Conference on Systems, Man and Cybernetics,* 565-569.

Sener, Z., & Karsak, E. E. (2008). A decision making approach based on fuzzy regression and fuzzy multiple objective programming for advanced manufacturing technology selection. *IEEE International Conference Industrial Engineering and Engineering Management,* 964-968.

Shang, J., & Sueyoshi, T. (1995). A unified framework for the selection of a flexible manufacturing system. *European Journal of Operational Research*, *85*, 297–315. doi:10.1016/0377-2217(94)00041-A

Shieh, J. I., Wu, H. H., & Liu, H. C. (2009). Applying a complexity-based Choquet integral to evaluate students' performance. *Expert Systems with Applications*, *36*(3), 5100–5106. doi:10.1016/j.eswa.2008.06.003

Small, M. H., & Chen, I. J. (1997). Economic and strategic justification of AMT inferences from industrial practices. *International Journal of Production Economics*, *49*, 65–75. doi:10.1016/S0925-5273(96)00120-X

Sugeno, M. (1974). *Theory of fuzzy integrals and its applications*. Unpublished doctoral dissertation. Tokyo Institute of Technology, Japan.

Talluri, S., Whiteside, M. M., & Seipel, S. J. (2000). A nonparametric stochastic procedure for FMS evaluation. *European Journal of Operational Research*, *124*(3), 529–538. doi:10.1016/S0377-2217(99)00188-5

Talluri, S., & Yoon, K. P. (2002). A cone-ratio DEA approach for AMT justification. *International Journal of Production Economics*, *66*(2), 119–129. doi:10.1016/S0925-5273(99)00123-1

Thakur, L. S., & Jain, V. K. (2008). Advanced manufacturing techniques and information technology adoption in India: A current perspective and some comparisons. *International Journal of Advanced Manufacturing Technology*, *36*, 618–631. doi:10.1007/s00170-006-0852-4

Tsai, H. H., & Lu, I. Y. (2006). The evaluation of service quality using generalized Choquet integral. *Information Sciences*, *176*(6), 640–663. doi:10.1016/j.ins.2005.01.015

Verter, V., & Dasci, A. (2002). The plant location and flexible technology acquisition problem. *European Journal of Operational Research*, *136*(2), 366–382. doi:10.1016/S0377-2217(01)00023-6

Weber, S. F. (1993). A modified analytic hierarchy process for automated manufacturing decisions. *Interfaces*, *23*(4), 75–84. doi:10.1287/inte.23.4.75

Yurdakul, M. (2004). Selection of computer-integrated manufacturing technologies using a combined analysis hierarchy process and goal programming model. *Robotics and Computer-integrated Manufacturing*, *20*, 329–340. doi:10.1016/j.rcim.2003.11.002

Yusuff, R. M., Yee, K. P., & Hashmi, M. S. J. (2001). A preliminary study on the potential use of the analytical hierarchical process (AHP) to predict advanced manufacturing technology (AMT) implementation. *Robotics and Computer-integrated Manufacturing*, *17*, 421–427. doi:10.1016/S0736-5845(01)00016-3

Zimmermann, H. J. (1991). *Fuzzy set theory and its applications*. Dordrecht, Germany: Kluwer Academic.

KEY TERMS AND DEFINITIONS

Advanced Manufacturing Technology (AMT): A modern method of production incorporating highly automated and sophisticated computerized design and operational systems.

Automated Storage and Retrieval System (AS/RS): Consists of a variety of computer-controlled methods for automatically placing and retrieving loads from specific storage locations.

Choquet Integral: A generalization of the Lebesgue integral. Choquet integral is a way of measuring the expected utility of an uncertain event.

Decision Making: Can be regarded as an outcome of mental processes leading to the selection of a course of action among several alternatives.

Fuzzy Logic: A form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise.

Fuzzy Sets: Are sets whose elements have degrees of membership and are extension of the classical notion of set.

Group Decision Making: Is decision making in groups consisting of multiple members.

Multi-Criteria Decision Making (MCDM): A discipline aimed at supporting decision makers who are faced with making numerous and conflicting evaluations.

This work was previously published in Technologies for Supporting Reasoning Communities and Collaborative Decision Making: Cooperative Approaches, edited by John Yearwood and Andrew Stranieri, pp. 193-212, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 61
Operator Assignment Decisions in a Highly Dynamic Cellular Environment

Gürsel A. Süer
Ohio University, USA

Omar Alhawari
Royal Hashemite Court, Jordan

ABSTRACT

Operators are assigned to operations in labor-intensive manufacturing cells using two assignment strategies: Max-Min and Max. The major concern is to see how these two approaches impact operators' skill levels and makespan values in a multi-period environment. The impact is discussed under chaotic environment where sudden changes in product mix with different operation times are applied, and also under non-chaotic environment where same product mix is run period after period. In this chapter, operators' skill levels are affected by learning and forgetting rates. The Max-Min strategy improved operators' skill levels more significantly than Max in this multi-period study; particularly in chaotic environment. This eventually led to improved makespan values under Max-Min strategy.

INTRODUCTION

Cellular manufacturing is considered as a collection of manufacturing cells that is dedicated to manufacture part families or assembly cells that are dedicated to process product families (see Askin & Standridge, 1993). The cellular manufacturing systems can be either machine-intensive or labor-intensive. In labor-intensive cells, it is easier to reconfigure cells when a product is ready to be processed. Moreover, moving equipment is much easier than it is in machine-intensive cells. Basically, in labor-intensive cells, most of the operations require light-weight, and small machines as well as equipment that require continuous operator attendance and involvement (Süer & Tummaluri, 2008). Labor-intensive manufacturing cells have been observed in apparel, jewelry manufactur-

DOI: 10.4018/978-1-4666-1945-6.ch061

ing, electromechanical assembly, sewing, shoe manufacturing, medical devices, and car seat manufacturing industries. The operator's role in machine-intensive cells is limited due to the presence of automatic machines. On the other hand, the operator has a key role in labor-intensive cells, and the number of operators and their assignment to operations has a great impact on the cell's production rate. In some cases, the number of operations is less than the number of operators. This creates the possibility that multiple operators are assigned to perform the same operation. It is important to control operator assignments; however, when the number of cells and the number of operators increase, keeping track of operator assignment becomes difficult.

In this chapter, concepts such as learning and forgetting rates are discussed to show how operator skill level varies from time to time; thus, the assignment decision is affected. Forgetting and learning rates affect the operator's skills and they are affected by their current skills. Learning takes place when the operator performs an operation continuously for a period of time, consequently, the operator will be more familiar with performing an operation. On the other hand, forgetting happens when the operator does not perform an operation in a number of consecutive periods. This chapter addresses both operator assignment and cell loading decisions. Operator assignment determines which operators are assigned to perform each task and cell loading identifies the products to be run in each cell.

The work undertaken in this chapter is an extension of work by Süer and Tummaluri (2008). The operator assignment can be made by using two different strategies; 1) Max, 2) Max-Min. Max considers only the current state of the operator skills for operator assignment to maximize output rate. On the other hand, Max-Min considers long-term effect of assignment decisions and attempts to develop more homogeneous work force without sacrificing output rate. This homogeneous work force may be more effective in dealing with drastic variations in demand and product mix in the long-term.

The objective of this chapter is to propose better mathematical models for operator assignment and also compare the performance of two major strategies, Max and Max-Min, in highly dynamic cellular environments. The main hypothesis is that Max-Min is a better strategy in operator assignment in the long-run. We want to show that long-term planning may help companies to better prepare their workforce for long-term operation than short-sided approach where only the immediate periods are considered. This approach is especially important in highly fluctuating demand environments and also in companies where product mix can quickly change. It is easier to implement such a strategy in companies where workforce is stable with low turnaround rate.

BACKGROUND

In the literature, some researchers addressed areas related to this subject such as cell loading, operator assignment, skills, learning and forgetting rate and product sequencing. Süer (1996) discussed, in his paper, the subject of optimal operator assignment and cell loading in labor-intensive manufacturing cells. He stated that the operator assignment to cells influences production rate that each cell can produce. He proposed a two-phase methodology. In phase 1, he generated operator assignments for alternative manpower levels by using a mixed integer mathematical model. In phase 2, he found the optimal manpower levels for each cell and optimal product assignment to cells.

Nembhard (2001) discussed a heuristic approach for assigning workers to task based on individual learning rate. Basically, he ran experiments based on two conditions: a long production run and a short production run. Results were interpreted and showed that the heuristic approach have an impact on overall productivity. Best results were found when workers learn more gradually.

Nembhard and Osothslip (2002) discussed, in their paper, the operator behavior in terms of learning and forgetting rates; particularly, in the case of performing complex tasks. A study was conducted at a textile manufacturing plant, where different manual sewing tasks were available. Data was collected by studying the behavior of each worker over a period of one year. They used a model of individual learning and forgetting rate, which was introduced first by Nembhard and Uzumeri (2000) in order to measure the productivity rate. This model was applied to each operator, and operator learning and forgetting parameters were considered as dependent variables, whereas task complexity was considered as independent variable. Results were captured and then statistical analysis was done to find if there is a relationship between the variability of learning and forgetting rates with task complexity. Results indicated task complexity significantly affects the variance of worker learning and forgetting rates. For higher task complexities, workers are more varied in their learning and forgetting rate than they are at lower task complexities. The impact of task complexities on worker learning and forgetting affects worker assignment and productivity. Slomp et al. (2005) discussed cross-training decisions in a cellular manufacturing environment. They wanted to minimize the load of the bottleneck worker. In their study, they presented an integer programming model to calculate which workers have to be trained for which machines. Based on this model, they discussed the trade-off between the operating costs of the manufacturing cell, the costs of cross-training, and the workload among workers, they showed that the connection between workers and machines is really important to form chaining and this produces an efficient cross-training situation. In this case, workload can be shifted from heavier loaded worker to less loaded worker. Labor flexibility is needed in these environments. Unbalanced load may give feelings of unfairness in a team. Bidanda et al. (2003) presented the importance of focusing not only on technical issues, but also human issues in cellular manufacturing environments. Technical issues include cell formation and cell design, whereas human issues involve such as worker assignment strategies, skill identification, training, communication, autonomy, reward system, conflict management, and teamwork. They conducted a survey to show the importance of human issues in cellular manufacturing. The number of participants in the survey was 40, and consists of workers, managers and academicians. They were asked to rank the human issues. Their response was analyzed. The results showed that three major human issues in cellular manufacturing are communication, teamwork and training. The degree of autonomy was found the least important among all. The reward system was in the middle. The assignment strategies were found significant among academicians. The skill identification was found significant among managers and academicians whereas conflict management is significant among workers.

Shirase et al. (2001) developed a system of distributed production system which consists of some cell groups. They discussed a dynamic operator assignment method. In this cooperative method, they considered that whenever any cell in a group of cells is unable to meet the due date of certain part, it has the option to ask for one operator as a support from other groups. Eventually, cooperation is taking place between cell groups until all due dates are met. They generalized the idea in which some disturbances in a production system can be treated by cooperation between subsystems. Fan and Gassmann (1997) discussed allocation functions between worker and machines could influence the performance of manufacturing cells over a long period of time considered as 15 months. They concluded that skills development and knowledge are really important for keeping long-term competitiveness. Allocation of functions has an impact on the long term performance, in which the long period will give more vision to make the work smooth through absorbing the complexity of the nature of manufacturing environment.

Wirojanagud et al. (2005) discussed a strategic way to model worker differences for workforce planning. Impact of individual differences on management decisions is considered and then discussed. A problem of job shop environment has been formulated in a form of a mixed integer programming model. The major concern was to identify number of workers who will be hired, fired and cross-trained at a minimum cost. Experiments are run and then results are analyzed. Workers differences are playing a major role in making decisions in manufacturing system environment.

Suksawat et al. (2005) discussed the concept of evaluating the skill levels of workers. They developed a skilled worker–based scheduling method based on the skill evaluation and genetic algorithm application. They focused on the objective of improving the production rate by considering workers' skill levels. Süer and Dagli (2005) discussed manpower decisions in their paper. They considered two issues in their paper; product sequencing in a cell and cell loading. For the first issue, they target to minimize intra-cell manpower transfers. A three-phase methodology is proposed, in which optimal manpower level for each operation is found by using a mathematical model, a matrix for manpower transfers between products is formed, and traveling salesman problem is solved. For the second issue, they aim to find the optimal assignment of products to cells. Their objective is to minimize makespan and number of machines. Cesani and Steudel (2005) presented a research concerning labor assignment strategies, and their impact on the cell performance in cellular manufacturing environment. The term labor flexibility was discussed and referred to the movement of operators between cells and inside the same cell. Labor assignments, such as dedicated assignment, in which an operator is assigned to one or more machines, shared assignment in which two operators or more are assigned to one or more machines, and combined in which an operator is assigned as dedicated and shared together. They made their discussions based on workload sharing, workload balancing and bottleneck operations. Experiments and simulation models were implemented and discussed. They concluded based on results, that the balance in the operators' workload and the level of machine sharing are important factors to determine cell performance and behavior. They also referred to the importance of cross-training issue in improving cell performance.

Mahdavia et al. (2010) developed an integer mathematical model to design cellular manufacturing systems. They consider a dynamic environment as well. Their model deals with worker assignment as well as dynamic configuration of the cellular system. The overall objective is to minimize the total cost of inventory holding and backorder costs, inter-cell material handling cost, machine and reconfiguration costs and hiring, firing and salary costs of workers.

Süer, Arikan, Babayigit (2008) and Süer, Arikan, Babayigit (2009) developed fuzzy mathematical models for cell loading in labor-intensive cells subject to manpower restrictions. Süer, Cosner and Patten (2009) developed various mathematical models for cell loading and product sequencing in manufacturing cells. Süer, Subramanian and Huang (2009) developed several heuristic procedures and mathematical models for cell loading and product sequencing in a shoe manufacturing company.

Süer and Tummaluri (2008) discussed the problem of operator assignment to operations in labor-intensive cells. The operators are assigned to operations in multi-period context considering their skill levels, forgetting and learning rates. They developed a three-phase hierarchical methodology to solve this problem. The first phase is generating alternative operator levels for each product using operation standard times. The second phase is determining cell loads and cell sizes using standard times. The third phase is assigning operators to operations. A mixed integer mathematical model is used in all phases. Two different strategies (Max and MaxMin) are proposed for solving the operator assignment in the third phase. The results showed that when using Max Strategy, lower makespan allocation values are obtained, whereas Max-Min improved the skill levels more regularly. The work undertaken in this paper is an extention of their work. They found that Max-Min is superior to Max Strategy in terms of improving operator skill levels; however, they could not show that Max-Min Strategy is better in minimizing makespan as a result of improved skill levels. Their work assumed a static cellular environment, in which there are no new products entering the system and no product is leaving the system (i.e. product mix remained the same throughout the study). They have classified labor skills into nine categories following normal distribution. Their work also showed that some non-bottleneck operations became bottleneck after assigning operators. The reason for that is that they have done initial operator assignment based on standard times. When they reflected the effect of skills on processing times, some non-bottleneck operations became bottleneck and adversely affected output rates. In some cases, operators had to be re-assigned to fix the problem.

PROBLEM STATEMENT

This chapter introduces several improvements to work by Süer and Tummaluri (2008) where, 1) Max and Max-Min strategies are compared in a highly dynamic environment where product mix changes. i.e., new products enter the system and some products leave the system. It is believed that this will show the benefits of Max-Min strategy better in terms of minimizing makespan, 2) number of skill levels are reduced to seven. It is believed that this is more practical and realistic approach than having nine skill levels, 3) skill-level based processing times are used directly during operator assignment process. This helps to avoid re-computing of bottleneck operation, output rate and re-assignment of operators.

Methodology

In this section, the methodology used is described in detail.

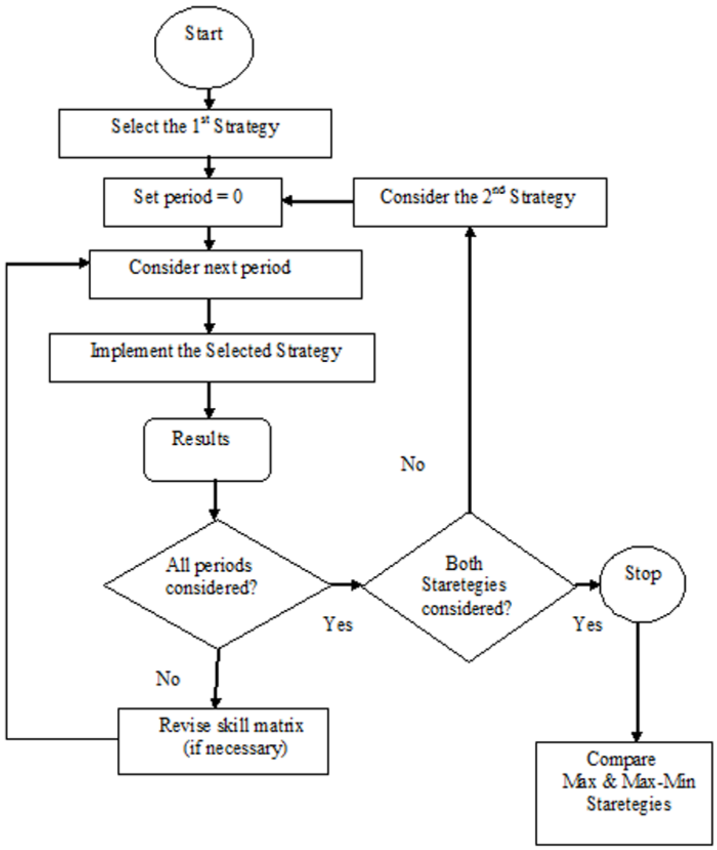
General Methodology

This study is carried out in a multi-period environment. The number of periods included in this study is 16 and each period represents a week. This allows us to see the impact of the strategies in the long-term. Based on the assignments made in the previous periods, an operator's skill level may be adjusted. Figure 1 includes the multi-period methodology in which the proposed approach is implemented and the results are captured. These results may affect the operator skill levels; hence, they need to be revised each period. Once all periods are considered for the first strategy, then the same procedure is applied for the second strategy, and finally the results are compared.

Overview of Strategies

The major phases of the Max strategy as shown in Figure 2a are: (1) find optimal operator assignment using an integer mathematical model. In this phase, the number of workers needed for each operation is determined for each product so that production rate is maximized with the available manpower level. (2) determine cell loads to minimize makespan by using an integer mathematical model. In this phase, decision about what cell to use to produce each product is made. (3) determine product sequence in each cell by using a simple scheduling rule, SPT (Shortest Processing Time). In this phase, the products in each cell are sequenced in the increasing order of processing time to minimize average flow time as well. The first three phases of the Max-Min Strategy is similar to the Max Strategy. However,

Figure 1. Multi-period methodology



two more phases are included in the Max-Min Strategy (see Figure 2b), and they are: (4) identify bottleneck and non-bottleneck operations for each product in the cell. The slowest operation (lowest output) for each product is identified as bottleneck operation. (5) re-assign low-skilled operators to non-bottleneck operations using the Min-skill principle. In this phase, the focus is on non-bottleneck operations and workers are assigned to operations where their skill level is not very high as long as it does not adversely affect the production rate of the cell. By doing this, we expect that operators with low skills will get a chance to perform these operations and eventually improve their skill levels.

Performance Measures Used

The performance measures used for each task are summarized in Table 1. Production rate measures the number of units manufactured per unit time. Makespan is defined as the maximum completion time of all jobs (Equation 1) and flowtime measures how long a job remains in the system (Equation 2).

$$MS=C_{\max} \quad (1)$$

$$f_i=c_i-r_i \quad (2)$$

where

c_i completion time of job i

Figure 2. The general overview of the strategies

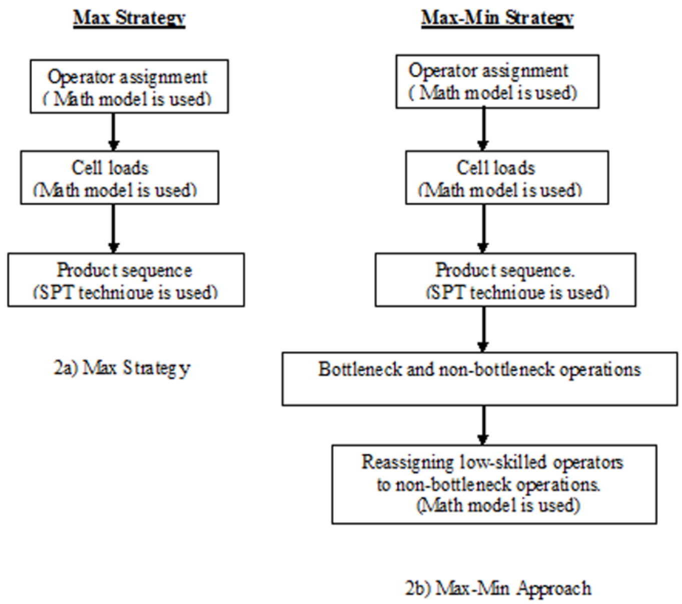


Table 1. Performance measures used for each task

Task	Max Strategy	Max-Min Strategy
Operator Assignment	Maximize Production Rate	Maximize Production Rate
Loading Cells	Minimize Makespan	Minimize Makespan
Product Sequencing	Minimize Average Flow Time	Minimize Average Flow Time
Re-assigning low-skilled operators	-----	Minimize Total Skills Without Violating Original Production Rate

f_i flowtime of job i
r_i ready time of job i

Skill Levels

In this study, each operator is assumed to have a skill level for each operation he performs. These skill levels follow the normal distribution (as shown in Figure 3), in which μ represents the mean value and σ represents the standard deviation. The skill levels are divided into seven categories and their corresponding probabilities are shown in Table 2. Level 4 represents the average skill, level 7 represents the best and level 1 is the worst. This is an assumption used in this study. Süer and Tummaluri (2008) also used a similar classification except they used 9 skills as opposed to 7 suggested here.

Operation Times

Operation times are calculated according to the operator skill levels. The standard processing time for each operation is considered to be the average time, hence, the operator with skill level 4 is considered to have the average operation time. Other skills below or over will follow the normal distribution. Table 3 provides an example for different skills of an operation with a standard deviation of 5% of the mean. The σ for operation

Figure 3. The normal distribution curve for skill levels

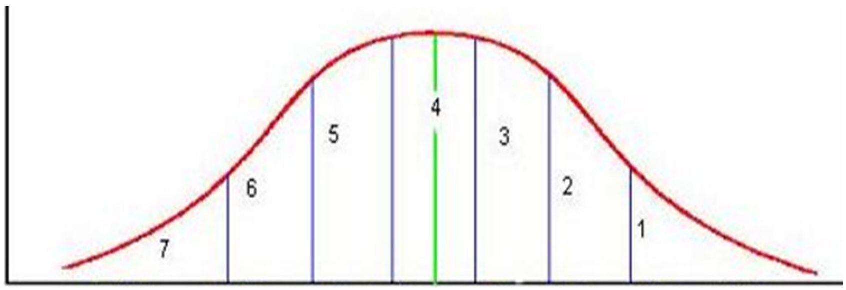


Table 2. The skill levels and probability

Skill Level	Time	Probability
1	μ +3σ	0.0062
2	μ+2 σ	0.0606
3	μ+σ	0.2417
4	μ	0.383
5	μ -σ	0.2417
6	μ-2 σ	0.0606
7	μ -3σ	0.0062

1 for product X1 is 0.0035 (=.07*.05). As the skill level decreases by 1 level (4 → 3), the operator skilled-based time increases to 0.0735 (=.07 + σ).

Learning and Forgetting Rates

In this study, a skill level is affected by learning and forgetting rates. An operator's skill level increases when he performs an operation for many consecutive periods. In the same manner, the length of interruption interval affects the skill levels adversely; hence, if an operator does not perform a certain operation for a number of consecutive periods, his skill level decreases. Table 4 shows the assumed number of required periods for improving or lowering skill level, and it also shows the probability for that skill to change. The probability values follow the notion that an operator who has been performing an operation for a long time will become more experienced operator, and it will take longer for that skill to deteriorate. As shown in the same table, a skill level can be improved from 1 to 2 with a probability of 0.7, if an operator keeps performing the same operation for 3 periods consecutively. On the other hand, improving skill level from 6 to 7 requires an operator to perform the operation for 6 periods consecutively (with a probability of 0.3). Meanwhile, an operator's skill level can deteriorate from 2 to 1 with a probability of 0.7, if he does not perform the operation for 4 periods consecutively. Similarly, if he does not perform the operation for 7 periods consecutively, his skill will decrease from 7 to 6 with a probability of 0.3. No empirical study has been done to validate these assumptions due to time restrictions. However, it is believed that these numbers reflect the relations between learning and forgetting rates of workers at different skill levels along with associated probabilities reasonably well. The operator skill matrix is revised on learning and forgetting rates at the end of every period.

Table 3. The operation times for different skill levels with σ = 5% of the mean

		Skills						
Mean	**Std. Dev.**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
0.07	0.035	.0595	0.063	0.0665	0.07	0.0735	0.077	0.0805

Table 4. The learning and forgetting rates

Learning				Forgetting			
Skill level		Periods	Prob.	Skill level		Periods	Prob.
From	To			From	To		
1	2	3	0.70	2	1	4	0.70
2	3	3	0.65	3	2	4	0.65
3	4	4	0.6	4	3	5	0.6
4	5	4	0.5	5	4	5	0.5
5	6	5	0.4	6	5	6	0.4
6	7	6	0.3	7	6	7	0.3

PROPOSED MATHEMATICAL MODELS

In this section, the proposed models are introduced.

Max Strategy

The Max Strategy uses the skill-based times to assign operators into operations such that maximum output is achieved. An integer mathematical model is used to assign operators. The mathematical model is formulated with the objective function of maximizing the production rate as shown in Equation (3). Equation (4) determines which operators have to be assigned to each operation. Equation (5) ensures that each operator is assigned to only one operation within a cell. Equation (6) shows that y_{kj} is a binary variable. The mathematical model is given below:

Objective Function:

$$MaxZ=R \quad (3)$$

Subject to:

$$\sum_{k\in fj}(a_{kj}y_{kj}) - R \geq 0 \quad j=1,2,3,\ldots,m \quad (4)$$

$$\sum_{j\in fk} y_{kj} = 1 \quad k=1,2,3,..,n \quad (5)$$

$$y_{kj}\in(0,1) \quad (6)$$

Indices:

- k Operator index
- j Operation index

Parameters:

- a_{kj} number of units operator *k* can process if assigned to operation *j*
- m number of operations in the cell
- f_k set of operations that operator *k* can perform
- f_j set of operators who can perform operation *j*
- n Number of operators

Decision variables:

- R production rate
- y_{kj} 1 if operator *k* is assigned to operation *j,* 0 otherwise

This assignment model is run for each product by using ILOG OPL software.

Max-Min Strategy

The Max-Min strategy also uses the skill-based times to find the optimal operator assignment to maximize the output. In this Strategy, first the slowest operation, bottleneck operation, is identified. The same integer mathematical model is used to assign operators and it is solved by using ILOG/OPL. However, another constraint is added to determine production rate for each operation as given in Equation (7).

$$R_j = \sum_{k \in fj} (a_{kj} y_{kj}) \text{ j=1,2,3..m} \quad (7)$$

Obviously, the lowest production rate operation is identified as the bottleneck operation as shown in Equation (8)

$$R_b = \min(R_j / j = 1,2,3,\ldots,m) \quad (8)$$

The Max-Min Strategy keeps the operators assigned to the bottleneck operation the same; however, it re-assigns other operators to non-bottleneck operations to minimize total skills such that the optimal output rate is maintained. A mathematical model is used where the objective function is to minimize the total skills for the remaining operators as given in equation (9). Equation (10) guarantees that the original production rate (optimal) is not violated. Equation (11) shows a constraint in which each operator is assigned to one operation within each cell. Equation (12) shows that y_{kj} is a binary variable.

The math model used is given below as:

Objective function:

$$MinZ = \sum_{k \in fj} \sum_{j \in fk} (s_{kj} y_{kj}) \quad (9)$$

Subject to:

$$\sum_{k \in fj} (a_{kj} y_{kj}) \geq R_b \text{ j=1,2,3,….m} \quad (10)$$

$$\sum_{j \in fk} y_{kj} = 1 \text{ k=1,2,3,…n} \quad (11)$$

$$y_{kj} \in (0,1) \quad (12)$$

where,

Parameters:

s_{kj} Skill level of operator *k* for operation *j*

R_b Production rate of bottleneck operation

The difference between these two strategies is illustrated in Figure 4 using a hypothetical case. Assume; that operation 2 is the bottleneck operation with operators 3 and 7 assigned to it as shown in Figure 4a and the output rate is 80 units/hr. The Max-Min Strategy shown in Figure 4b keeps the same operators in the bottleneck operation but re-assigns other operators to minimize skills without sacrificing the optimal output rate.

Cell Loading and Product Sequencing

Cell loading is the process of assigning products to cells. In this paper, a mathematical model is used to assign products to cells and the primary performance measure is to minimize makespan. Equation (13) shows the objective function, minimizing makespan. Equation (14) shows that the total processing time in each cell should be equal to or greater than the makespan. Equation (15) ensures that each product is assigned to a cell. Equation (16) shows the sign restriction.

Objective function

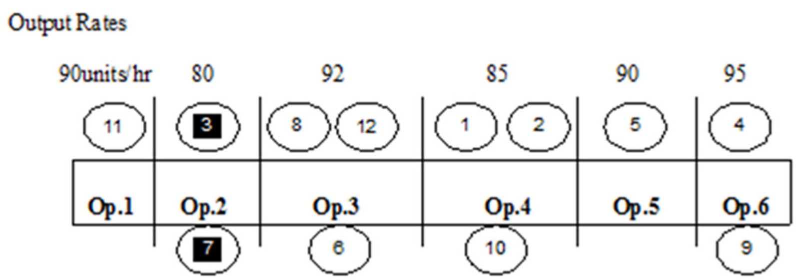
$$MinZ = MS \quad (13)$$

Subject to:

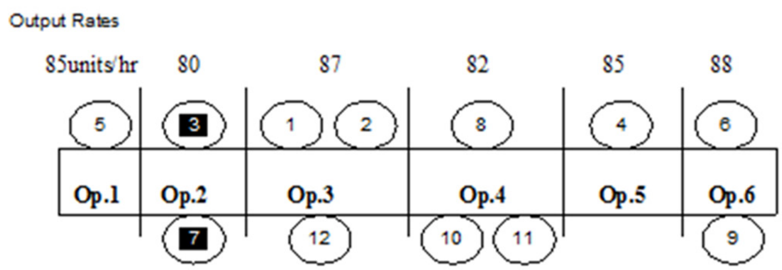
$$MS - \sum_{i=1}^{n} p_{ij} x_{ij} \geq 0 \text{ j=1,2,3,…,c} \quad (14)$$

$$\sum_{j=1}^{c} x_{ij} = 1 \text{ i=1,2,3,…,n} \quad (15)$$

Figure 4. Manpower assignment using both strategies



4a. The optimal operator assignment using the MAX Strategy



4b. The optimal operator assignment using the Max-Min Strategy

$x_{ij} \in (0,1)$ (16)

Where,

Indices:

c Number of cells

n Number of products

Parameter:

p_{ij} Processing time of product i in cell j

Decision Variable:

x_{ij} 1 if product i is assigned to cell j, 0 otherwise

Product sequencing usually comes after cell loading in which products are arranged in such a way, that a selected performance measure is completed. In this chapter, average flow time is considered as a secondary measure and it is minimized by using the shortest processing time technique (SPT). SPT rule orders jobs in the increasing order of processing times.

DATA USED IN EXPERIMENTS

In this section, the data used in experiments are given.

Standard Times

The standard times for operations correspond to a skill level of 4. The standard times for product groups X and Y are randomly generated from a random uniform distribution in the intervals shown in Table 5. These two product groups are used to create the dynamic environment mentioned earlier.

Product Demand

The product demand for all products is randomly generated from the uniform distribution in the interval of [2200, 8500] for all periods. Table 6 shows these demand values.

Table 5. Uniform distribution intervals for standard times for product groups X and Y

Product	Operations					
	Op.1	Op.2	Op.3	Op.4	Op.5	Op.6
X	[0.04-0.09]	[0.28-0.45]	[0.37-1.18]	[0.47-0.88]	[0.18-0.45]	[0.20-0.80]
Y	[1.10-1.30]	[1.20-1.40]	[0.09-0.15]	[0.04-0.09]	[1.00-1.20]	[0.04-0.08]

Table 6. Product demand and total demand for each period

Period	Product Demand										Total Demand
	1	2	3	4	5	6	7	8	9	10	
1	3500	7500	3400	2700	2200	4000	4500	2200	2300	3000	35300
2	3500	7500	3700	2900	2200	4300	4600	2200	2500	3000	36400
3	3700	4200	3700	3000	2400	4300	4600	2400	2500	3100	37200
4	3100	6500	3700	2750	2150	3500	4400	1900	2500	3100	33600
5	3100	3750	3700	2750	2400	3900	4200	1900	2500	3300	34500
6	4200	8000	3700	3000	2500	4300	4800	2400	2500	3100	38500
7	2800	6400	3700	2250	2150	3200	4400	1900	2300	3100	32200
8	2800	6300	3700	2250	2150	3200	4400	1900	2300	3100	31800
9	3100	7750	3700	2750	2400	3900	4200	1900	2500	3300	35500
10	3300	8450	3700	3050	2600	3900	4400	2000	2500	3500	37400
11	2900	6400	3700	2300	2150	3250	4400	1900	2300	3200	32500
12	3500	7600	3400	3200	2200	3500	4500	2200	2300	3000	35400
13	4100	7700	3500	2500	2400	4800	4000	2400	2500	3400	37300
14	2800	6800	3450	2500	2150	3200	4200	2100	2300	3100	32600
15	3600	6750	3700	3000	2400	3900	4200	1900	2750	3500	35700
16	3400	6900	3700	2900	2200	4300	4600	2200	2500	3600	36300

Operator Skills

The total number of operators included in the study is 30. Each operator is assumed capable of performing 3 operations. These operators are divided into two cells with 12 operators in cell 1 and 18 operators in cell 2. The initial skill matrix is established randomly by following probabilities given in Table 2. The initial skill matrix for all operators is given in Table 7.

EXPERIMENTS

Several experiments are conducted using 10 products of type X and 10 products of type Y. Each product requires 6 operations. The experiments performed are listed below:

1. Periods 1-14, no chaos
2. Periods 1-9, no chaos; Periods 10-11, chaos
3. Periods 1-14, no chaos; Periods 15-16 chaos
4. Periods 1-14, no chaos; Period 15 chaos with new standard deviation

Table 7. Initial operator skill matrix

Operator	Operations					
	Op1.	**Op.2**	**Op.3**	**Op.4**	**Op.5**	**Op.6**
1	5		5	4		
2		3	4	4		
3	3	5	3			
4				4	3	4
5	2	3			5	
6			5	6		4
7	4				4	3
8	4		4	4		
9		4			4	4
10	4		3	4		
1	4			1		3
12		3	4		5	
13	4	5		4		
1		5	5	5		
1	3		3		5	
16		3	3		5	
17	5			4		3
18	5	5		5		
1		5		5		4
20	5		3		6	
21	4			3		2
22		3	5		3	
23	3		4			5
24	4			5	3	
25			4	3		6
26		3	1			7
27	3			3	5	
28		4	2		4	
29	5		4			4
30	5				4	5

Experiment 1

This experiment includes runs using Max and Max-Min strategies starting from period 1 to period 14. The results are analyzed below:

Impact on Operator Skill Levels

It was found that Max-Min improves operator skill levels more significantly than Max did. Benefits come from the Max-Min assignment strategy in which it finds the optimal operator assignment to maximize the output. Later, this strategy keeps the operators assigned to the bottleneck operation the same but re-assigns the low skilled operators to non-bottleneck operations. This strategy allows an operator to perform an operation that he or she is not skilled at; hence his or her skill does not deteriorate and certainly improves. Tables 8 and 9 show the comparison of these two strategies at the end of period 14 in terms of average operator skill levels for cells 1 and 2, respectively. In cell 1, it was found that not only the average operator skill levels is greater than the initial average skill levels when Max-Min was used, but also is greater than operators average skill levels when Max was used. On the other hand, when Max was used, 5 operators had an average skill levels greater than their initial ones, 3 operators kept the

Table 8. Comparison of two strategies in cell 1 at the end of period 14

Cell 1		Average operator skill levels (at the end of Period 14)	
operator	**Initial average skill level**	**Max-Min Strategy**	**Max Strategy**
1	4.67	5.67	4.67
2	3.67	5.33	4.67
3	3.67	4.67	3
4	3.67	6	3.67
5	3.33	5.33	3.67
6	5	6.33	4.67
7	3.67	5.33	3.37
8	4	5	4.33
9	4	5.67	4.33
10	3.67	5.33	3.67
11	2.5	5.67	3.67
12	4	5.33	3.67

Table 9. Comparison of two strategies in cell 2 at the end of period 14

Cell 2		Average operator skill levels (at the end of Period 14)	
operator	**Initial average skill level**	**Max-Min Strategy**	**Max Strategy**
13	4.33	5.33	4.33
14	5	7	6
15	3.67	6.33	5.67
16	3.67	6.33	5.33
17	4	5.33	4.67
18	5	6	5.33
19	4.67	5.67	5
20	4.67	6	5
21	3	5.67	5.67
22	3.67	5.33	3.33
23	4	6	5
24	4	5	3.67
25	4.33	6.33	5.33
26	3.67	4	3
27	3.67	6	5.33
28	3.33	5.33	5.33
29	4.33	5.33	4.67
30	4.67	6	6.67

same average and 4 operators were found having lower average skill levels.

In cell 2, it was found that the average operator skill level is greater than the initial average skill levels when Max-Min was used. However, 15 operators had greater average skill levels than when Max was used, but 2 operators had the same average skill levels and 1 operator had lower average skill levels. As to Max strategy, 14 operators had an average skill levels greater than their initial ones, 1 operator kept the same average and 3 operators were found having lower average skill levels than their initial ones.

Impact on Operations

Max-Min Strategy improved operator skill levels on each operation more apparently than Max did. Table 10 shows average skill levels on each operation. It was found that operation 1 deteriorated when Max was used; however average operator skills on all operations were greater when Max-Min was used.

Impact on Makespan

The results have shown that both strategies were tied in terms of makespan for the first 11 periods. Starting period 12, Max-Min performed better in minimizing makespan by 0.2%, 0.2% and 0.1% compared to Max approach in periods 12, 13 and 14, respectively. Table 11 shows the comparison of these two strategies in terms of makespan from period 1 to period 14.

Experiment 2

In this section, chaos is applied in periods 10 and 11 to create a big shock in the system to see how both strategies behave in terms of makespan. This is accomplished by introducing new products with very different processing times. The results have shown that when 5 products of type Y entered the system and 5 products of type X left the system, Max-Min gave better results (with 0.4% reduction

Table 10. Average skill levels on each operation

		Average operator skill levels (at the end of period 14)	
Operation	**Initial average skill level**	**Max-Min Strategy**	**Max Strategy**
Op.1	4	4.39	3.44
Op.2	3.92	5.23	4.23
Op.3	3.65	6	4.89
Op.4	4	6.25	5.63
Op.5	4.31	6.15	4.62
Op.6	4.25	5.92	4.62

Table 11. The makespan from period 1 to period 14

period	Max	Max-Min	Difference(%)
1	58.07	58.07	0.0%
2	59.84	59.84	0.0%
3	61	61	0.0%
4	53.13	53.13	0.0%
5	55.92	55.92	0.0%
6	60.07	60.07	0.0%
7	50.23	50.23	0.0%
8	49.684	49.684	0.0%
9	55.76	55.76	0.0%
10	57.9	57.9	0.0%
11	50.64	50.64	0.0%
12	53.19	53.1	0.2%
13	56.25	56.15	0.2%
14	49.04	48.99	0.1%

in makespan) in period 10. In period 11, the entire set of products of type Y is manufactured (no product X in the system). Max-Min still worked better (with 0.4% improvement in makespan). Table 12 shows the results of makespan using these two strategies in periods 10 and 11. Table 13 shows another chaotic scenario by entering all products of type Y and releasing all products of type X in period 10. In this case, the improvement with Max-Min over Max was 0.7%.

Table 12. Impact on makespan under chaos in periods 10 and 11

	Max-Min	Max	Difference (%)
Period 10 (5 products)	71.15	71.43	0.4%
Period 11 (5+5 products)	81.68	82.02	0.4%

Table 13. Impact on makespan under chaos in period 10

	Max-Min	Max	Difference (%)
Period 10 (10 products)	94.3	94.92	0.7%

Experiment 3

The chaos was applied in periods 15 and 16 as we did in experiment 2. The results have shown that when 5 products of type Y entered the system and 5 products of type X left the system, Max-Min gave better results (with 1.7% reduction in makespan) in period 15. In period 16, the entire set of products of type Y is manufactured (no product X in the system). Max-Min still worked better (with 1.2% improvement in makespan). Table 14 shows the results of makespan using these two strategies in periods 15 and 16. Table 15 shows another chaotic scenario by entering all products of type Y and releasing all products of type X in period 15. In this case, the improvement with Max-Min over Max was 2.2%.

Experiment 4

In this phase, we wanted to show that the improvement in operators' skill levels should have better impact than what happened in previous experiments. We changed the standard deviation from (0.05μ) to (0.2μ) to extend the gap of operator-operation times, and then we applied these new times in period 15 by entering all products of type Y in the system. Results have shown that there was better gain in terms of minimizing makespan

Table 14. Impact on makespan under chaos in periods 15 and 16

	Max-Min	Max	Difference (%)
period 15 (5 products)	66.48	67.6	1.7%
period 16 (5+5 products)	88.6	89.7	1.2%

Table 15. Impact on makespan under chaos in period 15

	Max-Min	Max	Difference (%)
Period 15 (10 products)	86.81	89.01	2.2%

Table 16. Impact on makespan under chaos in period 15

	Max-Min	Max	Difference (%)
period 15 (10 products) (SD = 0.2)	55.23	58.02	5%

and total time. Max-Min performed better by 5% than Max approach. Table 16 show these results.

CONCLUSION AND FUTURE WORK

In this chapter, operators were assigned to operations to maximize the production rate using two assignment strategies: Max-Min and Max. The major concern is to see how these two approaches impact operators' skill levels, as well as their impact on makespan values. The impact is discussed under a chaotic environment where sudden changes in product mix with different operation times are applied and under a non-chaotic environment where the same product mix is run period after period. In this study, a skill level is affected by learning and forgetting rates. An operator's skill level increases when he performs an operation for many consecutive periods; on the other hand, if an operator does not perform a certain operation for a number of consecutive periods, his skill level decreases. Max-Min did improve operators' overall skill levels more significantly than Max in multi-periods. This is due to the fact that Max-Min does not only assign operators to maximize production rate, but also it re-assigns operators to operations where they are not very skilled; thus, operator's skill levels continue to improve. On the other hand, Max is only assigning operators to operations to maximize production rate. Moreover, the previous work assumed a stable cellular environment, in which they assumed that there are no new products entering the system. Thus, in this paper, we introduced a highly dynamic cellular environment, in which new products with different processing times entered the system and some of the existing products left the system. Max-Min acts well under chaotic environment because it increases operators' skill levels well enough to face the shock applied, where the shock contains products with new processing times and this requires different manpower allocation. We also concluded that the standard deviation used in operator time matrix is an important factor for helping Max-Min approach to expand the gain in terms of minimizing makespan and total time. A standard deviation of 5% of the mean is used in operator matrix for experiments 1 and 2. In period 15, when we replaced the whole set of products of type X with products of type Y, we found the highest gain that shows Max-Min is better in minimizing makespan among all periods. This gain was captured using a standard deviation of (0.05μ); however, when we used a standard deviation of (0.2μ), we found that the gain is higher than using (0.05μ). A possible extension to this work is to take manpower level decision for each cell as a variable as opposed to using fixed manpower levels. A further expansion would be to allow operators shift from one cell to the next as opposed to fixing them to the same cell all the time. Obviously, this increases flexibility in assigning operators to cells. However, this increases computational complexity as well.

REFERENCES

Askin, R., & Standridge, C. (1993). *Modeling and analysis of manufacturing systems*. New York, NY: John Wiley & Sons, Inc.

Bhaskar, K., & Srinvasan, G. (1997). Static and dynamic operator allocation problems in cellular manufacturing systems. *International Journal of Production Research*, *35*(12), 3467–3481. doi:10.1080/002075497194192

Bidanda, B., Ariyawongrat, P., Needy, K. L., Norman, B. A., & Tharmmaphornphilas, W. (2005). Human related issues in manufacturing cell design, implementation, and operation: A review and survey. *Computers & Industrial Engineering*, *48*(3), 507–523. doi:10.1016/j.cie.2003.03.002

Cesani, V. I., & Steudel, H. J. (2005). A study of labor assignment flexibility in cellular manufacturing systems. *Computers & Industrial Engineering*, *48*(3), 571–591. doi:10.1016/j.cie.2003.04.001

Fan, S., & Gassmann, R. (1997). The affects of allocation of functions on the long-term performance of manufacturing cells—A case study. *Human Factors and Ergonomics in Manufacturing*, *7*(2), 125–152. doi:10.1002/(SICI)1520-6564(199721)7:2<125::AID-HFM5>3.0.CO;2-5

Hyer, N. L. (1984). Management's guide to group technology. In Hyer, N. L. (Ed.), *Group technology at work* (pp. 26–27). Dearborn, MI: SME.

Mahdavia, I., Aalaei, A., Paydar, M. M., & Solimanpur, M. (2010). Designing a mathematical model for dynamic cellular manufacturing systems considering production planning and worker assignment. *Computers & Mathematics with Applications (Oxford, England)*, *60*(4), 1014–1025. doi:10.1016/j.camwa.2010.03.044

Nembhard, D. A. (2001). Heuristic approach for assigning workers to tasks based on individual learning rates. *International Journal of Production Research*, *39*, 1955–1968. doi:10.1080/00207540110036696

Nembhard, D. A., & Osothsilp, N. (2002). Task complexity effects on between-individual learning/forgetting variability. *International Journal of Industrial Ergonomics*, *29*(5), 297–306. doi:10.1016/S0169-8141(01)00070-1

Nembhard, D. A., & Uzumeri, M. V. (2000). Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics*, *25*, 315–326. doi:10.1016/S0169-8141(99)00021-9

Shirase, K., Wakamatsu, H., Tsumaya, A., & Arai, E. (2001). *Dynamic co-operative scheduling based on HLA.* Retrieved September 10, 2007, from http://www.springerlink.com/content/r116462352k35gt0/fulltext.pdf

Slomp, J., Bokhorst, J. A., & Molleman, E. (2005). Cross-training in a cellular manufacturing environment. *Computers & Industrial Engineering*, *48*(3), 609–624. doi:10.1016/j.cie.2003.03.004

Süer, G. A. (1996). Optimal operator assignment and cell loading in labor-intensive manufacturing cells. *Computers & Industrial Engineering*, *31*(1-2), 155–158. doi:10.1016/0360-8352(96)00101-5

Süer, G. A., Arikan, F., & Babayigit, C. (2008). Bi-objective cell loading problem with non-zero setup times with fuzzy aspiration levels in labour-intensive manufacturing cells. *International Journal of Production Research*, *46*(2), 371–404. doi:10.1080/00207540601138460

Süer, G. A., Arikan, F., & Babayigit, C. (2009). Effects of different fuzzy operators on fuzzy bi-objective cell loading problem in labor-intensive manufacturing cells. *Computers & Industrial Engineering*, *56*, 476–488. doi:10.1016/j.cie.2008.02.001

Süer, G. A., Cosner, J., & Patten, A. (2009). Models for cell loading and product sequencing in labor-intensive cells. *Computers & Industrial Engineering*, *56*, 97–105. doi:10.1016/j.cie.2008.04.002

Süer, G. A., & Dagli, C. (2005). Intra-cell manpower transfers and cell loading in labor-intensive manufacturing cells. *Computers & Industrial Engineering*, *48*(3), 643–655. doi:10.1016/j.cie.2003.03.006

Süer, G. A., Subramanian, A., & Huang, J. (2009). Heuristic procedures and mathematical models for cell loading and scheduling in a shoe manufacturing company. *Computers & Industrial Engineering*, *56*, 462–475. doi:10.1016/j.cie.2008.10.008

Süer, G. A., & Tummaluri, R. (2008). Multi period operator assignment considering skills, learning and forgetting in labour-intensive cells. *International Journal of Production Research*, *2*, 1–25.

Suksawat, B., Hilraokai, H., & Ihara, T. (2005). *A new approach manufacturing cell scheduling based on skill-based manufacturing integrated to genetic algorithm*. Retrieved October 15, 2007, from http://www.springerlink.com/content/q885171535723r5q/fulltext.pdf

Wirojanagud, P., Gel, S., Fowler, J., & Cardy, J. (2005). *Modeling inherent worker differences for workforce planning*. Retrieved September 10, 2007, from http://ie.fulton.asu.edu/files/shared/workingpapers/Wirojanagud_Gel_Fowler.pdf.

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 258-276, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 62
Capacity Sharing Issue in an Electronic Co-Opetitive Network:
A Simulative Approach

Paolo Renna
University of Basilicata, Italy

Pierluigi Argoneto
University of Basilicata, Italy

ABSTRACT

In recent years, manufacturing companies have entered a new era in which all manufacturing enterprises must compete in a global economy. To stay competitive, companies must use production systems that only produce their goods with high productivity, but also allow rapid response to market changes and customers' needs. The emerging new paradigm of inter-firm relations involving both cooperative and competitive elements, called co-opetition, seems well face this issue. The chapter proposes a multi agent architecture to support different coordination policy in an electronic co-opetitive network in which plants are willing to exchange productive capacity. An innovative approach based on cooperative game theory is proposed in this research and its performance is compared with the prevalent negotiation approach. A discrete event simulation environment has been developed in order to evaluate the related performances. The case in which no relation exists among plants has been considered as a benchmark. The obtained results show that the proposed approach outperforms the negotiation mechanism form many point of view.

DOI: 10.4018/978-1-4666-1945-6.ch062

INTRODUCTION

In this era of downsizing and outsourcing, where the business landscape is changing rapidly, the management of the external relations of the firm and the combining of this with internal operations represent a new problem and opportunity. Increasingly firms are recognizing the development and management of relationships both directly and indirectly with other members of the value producing system, conscious of the fact that they cannot survive and prosper solely through their own individual efforts: each firm's performance depends in important ways on the activities and performance of others and on the nature and quality of the direct and indirect relations as firm develops with these counterparts. Inter-firm relations involve a mix of cooperative and competitive elements. Firms simultaneously cooperate to expand the total amount of rewards and resources available to them and compete over the means to do this and over the division of rewards and resources. They compete to develop cooperative relations with counterparts such as customers and suppliers that are beneficial in order to create competitive advantages in creating value for final customers. Matching competition and cooperation (hereafter C&C) is a great and hard challenge for firms in this globalization era. The novelty of the topic proves the importance and the absolute relevance of research efforts aiming at providing a better understanding of this new organizational issue. Although some works are present in literature, identifying and providing evidence of the reasons of both competitive and cooperative relationships among firms, analyzing keys of success and causes of failures of these strategies are still open questions in both conceptual and empirical literature. On the other hand, determining strategies in these situations and studying their beneficial or detrimental effects on firms, consumers, suppliers and social welfare, designing proper mechanisms and providing incentives are absolutely open issues. When a rapport among firms includes elements of both cooperation and competition, i.e. these firms can compete and cooperate simultaneously, the relationship is called co-opetition. In order to better understand the reasons for assuming co-opetitive behavior, we should first reflect on the concepts of C&C. Competition occurs when many firms are producing the same or related products and are fighting for consumers, suppliers and/or resources, excluding collaboration with other companies. In pure competition the boundaries between competitors are sharp and distinct. Conversely, in pure cooperation there are frequent exchanges among partners, including business, information and social exchange. Inter-firm cooperation produces strong ties that cross the boundaries between firms in order to share complementary capabilities, assets and interests. The functioning of cooperative relationships may be regulated by formal contracts and/or informal agreements built upon social relations and the development of trust. It should be noted that cooperation and competition between firms are neither spontaneous nor exogenous, but are actions that depend upon the contextual conditions, i.e. the determinants of their level evolve over time, thus it is likely that the level of C&C between organizations will undergo variations (Luo, 2005). Some of the various determinants of competition are: a high level of overlapping in the market, a slight differentiation between the products sold by different firms, a lack of entry barriers to a market and a strong contractual position enjoyed by clients and/or suppliers. Instead, some determinants of cooperation among firms are: an inability to achieve their goals with their own resources, a need to share the risks of a new initiative and an improvement in efficiency through the sharing of activities to obtain economies of scale. For all that reasons, various hybrid forms of co-opetitive relationships occur between the extreme forms of pure inter-organizational competition and pure inter-organizational cooperation. Following the seminal work of Brandenburger and Nalebuff (1996), a number of articles regarding the private sector argued that C&C among firms cannot be considered mutually exclusive (Bengtsson and Kock, 1999; Luo, 2004; Oliver, 2004) but, in spite of authors' suggestion few of them have used game theory to model situations in which both competitive and cooperative behaviors arise. Game theory makes it possible to move beyond overly simple ideas of C&C to reach a vision of co-opetition more suited to topical needs. This theory is a branch of mathematics that is concerned with the actions of individuals who are conscious that their actions affect each other. It is divided into two branches, called the non-cooperative and cooperative ones, differing in how they formalise interdependence among the players. In the non-cooperative theory, a game is a detailed model of all the moves available to the players. By contrast, the cooperative theory abstracts away from this level of detail, and describes only the outcomes that results when the players come together in different combinations. Though standard, the terms non-cooperative and cooperative game theory are perhaps unfortunate. They might suggest that there is no place for cooperation in the former and no place for conflict or competition in the latter. In fact, neither is the case: especially in business field the real value of game theory comes when both approaches are put into practice. It, therefore, could be used to study settings in which co-opetition might be used, determine strategies of co-opetition and analyze their effects on outcome in order to evaluate possible advantages and/or disadvantages from single firm, other stakeholders and the whole system. For example, consider a situation where a number of firms (players) are connected in some network relationship. The applications could be quite wide and varied, ranging from friendships and social relationships, to communicating information about job openings, to business partnerships, to international trade agreements and political alliances. What is in common to these situations is that the way in which players are connected to each other is important in determining the total productivity or generated value. How the total productive value are allocated or transferred among players turns out to be important not only in terms of fairness considerations, but also because it determines players' incentives to form various networks. Particularly the generated value by players depends not only on their identities but also on how they are connected to each other. That is, alternative network structures (e.g., communication lines, alliances, friendships, etc.) connecting the same set of players might lead to very different costs and benefits. Thus, in many situations it will be important to account for network structure and not just coalition structure. Myerson (1977) made a seminal contribution in adapting the cooperative game theory structure to accommodate information about the network connecting players: once the network is fixed its role is simply to define which coalitions can function. The feasible coalitions are the ones whose members can freely communicate via the given network (so that any two players in the coalition are path connected in the network via players in the coalition). In a sense, each network structure and characteristic function (indicating how much value a given coalition can generate) induces a particular cooperative game.

In the last two decades, there have been many relevant examples of co-opetition situations. In the 90s GM and Ford, the major American carmakers, established an e-procurement platform for procuring basic components. The joint venture between Toyota and PSA Citroen-Peugeot, established in 2002, is another very relevant example of co-opetition in automobile industry. The two companies agreed in building a common plant in Czech Republic and using common components for the production of three new separately-owned city cars. In Italy, in 2002, the two biggest motorcycle companies, Aprilia and Piaggio made an alliance for joint-procurement, though competing in the final market. In ICT industry, the Simian joint-venture is among the main mobile wireless telephones manufacturers in the world, Nokia, Ericsson, Panasonic, Samsung, Siemens AG, and the leading company in the mobile digital computing, Psion, which, however, has sold its own shares in 2004. The phenomenon of co-opetition in R&D activities and co-promotion is also very common in pharmaceutical and biotechnology industry. Other straightforward examples of co-opetition can be found in several other industries, such as furniture, tourism, healthcare and retailing. Moreover, all around the world industrial districts and both local or national consortia and co-ops can be considered examples of co-opetition of medium and small businesses in several industries, such textile and clothing, agri-food, hardware and retailing. Also, in the automotive industry, web-based technology opens up opportunity of cooperation among small and medium firms. For instance, Supply On is a successful provider of supply and engineering services founded by suppliers competing in the same market (Meder, 2005). Finally, an interesting empirical research from Quintana-Garcia and Benavides-Velasco (2004) provides evidence of co-opetition in new product development among European small and medium firms in biotechnology industry. Through cooperation in one or more activities, small firms can get same advantages as big companies, however, because of proximity, they inevitably keep competing each others in other activities. The contribution of our chapter is threefold:

- to apply concepts deriving from co-opetitions in an electronic network of geographically distributed plants willing to exchange their productive capacity;
- to evaluate different coordination policy, one based on negotiation and one based on cooperative game theory;
- to simulate the proposed approach by a Multi Agent System able to support the network.

The rest of the chapter is structured as follows: section 2 presents the motivations to develop a network; section 3 provides a general description of the network typology while the co-opetitive approach is introduced in section 4. An overview of the literature is presented in section 5. In section 6 is described the problem context then the Multi Agent Architecture is illustrated in section 7. In section 8 the coordination mechanism proposed are described. The developed simulation environment and simulation results are respectively presented in sections 9 and 10. Finally, conclusions and further research paths are withdrawn in section 11.

WHY NETWORKS ARE NECESSARY

Survival, performance and development of the entrepreneurial firm are at the heart of entrepreneurship research (Shane and Venkataraman, 2000). Entrepreneurial firms are characterized by a lack of internal resources and other start-up handicaps, as expressed in the theoretical constructs of liability of newness and liability of smallness. The strategic use of external resources through inter-firm networks in many different industries (Jarillo, 1989), that are often embedded in regional clusters (Boari and Lipparini, 2000; Lechner and Dowling, 2000) is regarded as one effective means to overcome these liabilities. In this context, inter-firm networks are considered an important model of organization development to enable an entrepreneurial firm to grow and survive (Freel, 2000). The image of atomistic actors competing for profits against each other in an impersonal marketplace is increasingly inadequate (Granovetter, 1985; Gulati, 1998; Galaskiewicz and Zaheer, 1999). Such networks encompass a firm's set of relationships, both horizontal and vertical, with other organizations—be they suppliers, customers, competitors, or other entities— including relationships across industries and countries. These strategic networks are composed of inter-organizational ties that are enduring and include strategic alliances, joint ventures, long-term buyer-supplier partnerships and a host of similar ties. Gulati et Al. (2000) highlight the importance of considering the role of strategic networks resulting from inter-firm ties for fundamental issues in strategy research. Anand and Khanna (2000) provide a unique empirical assessment of the presence of experience benefits to firms from entering frequent alliances. Using a comprehensive data set on abnormal returns to alliance announcements they separate out alliances by contract type and find significant differences in the extent of learning, as measured by abnormal returns, across contract type. They look further within each type and also observe differences in learning effects within joint ventures depending upon the scope of activities within the alliance. They conclude that learning effects are more significant when there is contractual ambiguity. Doz et Al. (2000) examine the creation of R&D firms consortia to uncover the formation processes underlying the formation of cooperative interfirm networks. They identify an emergent process by which environmental changes and shared views among participants facilitate network formation, and also an engineered process in which a triggering entity actually recruits potential members. Together, the identification and elucidation of these two processes provide us with insights into the formation of networks. Baum et Al. (2000) utilize a unique data set on biotech start-ups to examine the role of start-ups' network composition on their innovative performance. Their findings suggest that the structure of their network of ties and the identity of their partners can have a significant influence on their performance. They use these empirical findings to reflect on some implications that may follow for managers of startups. Ahuja (2000) draws on the resource-based view of the firm to study alliance formation. He flags an important issue when he argues that the proclivity of firms to form strategic alliances depends both on alliance opportunities and on the firms' own inducements and resource endowments. Alliance formation thus results from a match between the value-creating resources of the focal firm and those available from other firms in its environment. He identifies three forms of capital resources (technological, commercial and social) and shows that they each contribute to alliance formation. A combination of technical and commercial capital in a firm, however, reduces the likelihood of allying, since such firms are already well endowed and have less need to ally. Rowley, Behrens and Krackhardt (2000) empirically assess the interplay between relational and structural embeddedness for firm performance. While the former highlights the role of the relationships firms are in, they focus on the overall relational structure surrounding the firm. Using data on interfirm strategic alliances from the semiconductor and steel industries they find that the role of both forms of embeddedness on firm performance is influenced by the industry context. Afuah (2000) explains firm-level competitive advantage by explicitly invoking the capabilities available to a firm from its network of supplier and customer firms (co-opetitors). He shows how a firm's performance is lowered in the face of technological change when a firm's network members' capabilities are made obsolete - a contrast from research which has focused on the capabilities of the firm itself in explaining performance. Kogut (2000) theoretically highlights two ways in which a firm's performance can be influenced by the network in which it is embedded. The first is the value that a focal firm's network of ties has for the range and quality of information it receives. This potential information advantage has been widely recognized. What has received less attention, Kogut argues, are the potential benefits that arise from being a part of a dynamic and evolving network that coordinates activities among specialized producers. By highlighting the importance of generative rules that shape the coordination among the members of a network over time.

NETWORK TYPOLOGIES

Generally, research on inter-firm networks has focused on the role of the entrepreneur in network building, on the initial size of an entrepreneurial firm's network in regard to firm performance, or on structural characteristics of networks. Different types of relations define different types of networks even if the same units are connected, that means that firms can be involved in different types of networks in different development phases. All this value-added networks go beyond economic relationships and includes:

- **social networks:** relationships with other firms based on strong personal relationships with individuals such as friends, relatives, long-standing colleagues that became friends before foundation, and so forth;
- **reputational networks:** made up of partner firms that are market leaders, or highly regarded firms or individuals, and where one of the main objectives in entering into this relationship is to increase the entrepreneurial firm's credibility;
- **marketing information networks:** relationships that allow for the flow of market information through distinct relationships with other individuals/firms;
- **co-opetitive networks:** relationships with direct competitors;
- **cooperative technology networks:** technology alliances involving joint technology development or innovation projects.

Moreover, network research distinguishes three components: network content, network structure and network governance in order to explain the role of networks in firm performance. According to the theory of structural embeddedness, network structure and a firm's network position are considered to be both opportunities and constraints. Favorable positions are regarded as network resources; over-embeddedness, however, can lead to inability to act. A rich literature suggests that networks are a particular governance form in which the development of trust plays a major role in influencing resource exchange and costs compared to market coordination or integration of activities. In this sense, inter-firm networks constitute a third way of organizing the business, which is neither by markets nor by hierarchies. Recently, it was shown that specific kinds of relations (network content) are more important in a different economic context. Research on firm networks as a mode of transaction governance, the role of strong and weak ties, and the analysis of structural properties of networks has produced important insights. First of all, many studies have shown that the entrepreneurs' personal and social networks probably are the most important strategic resources for the firms' start-up (Ardichvili et al., 2003). Organizational networks or inter-firm networks are relations between organizations that can have various functions, also called sub-networks according to their relational content. The merging of personal and organizational networks seems to be a common feature of young firms. Founders and the firm are inseparable at startup. As the firm grows, the founders' personal networks and firm networks merge (Cooper, 2002). These merged networks can be considered an organizational form. Initially, it seems that networking for entrepreneurial firms is based on pre-existing relationships, which become more complex over time by having different functions and being more socially embedded: social relations are transformed into socio-economic and, finally, into more complex relations. Previous research has shown that the overall network structure changes from an unplanned to a planned and, finally, a structured network. Once structured, however, it seems that both over-embeddedness and a firm's limited relational capability, i.e. the capability to establish, maintain, and develop relationships (Lechner and Dowling, 2003), poses a potential barrier to growth. However, this research does not explain which network types are most important to manage or how firms should overcome these growth barriers. Research on network types and performance has focused on pre-start-up activities in order to explain nascent entrepreneurship (i.e. personal networks have been analyzed at or prior to the creation of the company in terms of size and networking activity). Context-specific research investigated the positive or negative role of social networks at start-up, but which types of network matter most for firm performance have not been studied extensively. Some research was conducted on the role of single network types such as reputational and cooperative technology networks, but little is known for example about the role of co-opetition networks. Lechner and Dowling (2003) proposed a network development model based on varying network types. Based on case study research in a German Information Technology cluster, they identified that firms use relationships for a variety of purposes and that every firm has an individual relational mix. They argued that the relational mix (i.e., the different types of networks) changes over time in order to enable firm growth. Finally, they proposed a four-phase development model of entrepreneurial firms. In phase 1, firms seek to overcome liability of newness by basing the development of the network mainly on social (understood as strong ties such as family and friends) and reputational networks (relationships with prominent firms that can lend the young firm reputation). While the relative importance of social and reputational networks decreases with the firms' development, electronic co-opetitive networks (i.e., cooperation with competitors) increase over time. In phase 2, firms use marketing information and electronic co-opetitive networks to overcome the usual period of unstable sales growth; in phase 3, co-opetition remains a relevant issue but cooperative technology networks are most important. Finally, in phase 4, firm growth is limited by path-dependent relational capability that eventually reaches its limits and leads to the reconfiguration of a more stable network by introducing hierarchic levels within the network or to the integration of activities previously performed outside the firm. This qualitative study seems to be one of the few where different network types were used to understand the role of networks at and beyond foundation, but their impact on the performance of the entrepreneurial firm have not been tested empirically with larger samples.

WHY WE ARE INTERESTED TO THE ELECTRONIC CO-OPETITIVE NETWORK

Electronic co-opetitive networks involve relationships with direct competitors. The management literature generally considers industries to be collections of firms bound together by rivalry, therefore questioning the value of relationships with competitors (Dollinger, 1985). Firms can use competitors as subcontractors in times where the firm has temporarily reached full capacity. This cooperative behavior, especially with regional competitors, will increase the likelihood of the favor being returned. Overall, relationships with competitors can give access to temporarily needed resources or lead to the temporary pooling of resources, which should positively influence firm performance especially in the years after foundation, when sales tend to grow discontinuously (Lechner and Dowling, 2003). While it has been argued that electronic co-opetitive networks at foundation might be harmful because such relationships could lead to the disclosure of competitive information, lack of co-opetition can also constrain firm development in the years following foundation. Entrepreneurial firms that view competitors not only as pure rivals but also as a potential resource should therefore be more successful. Building on the theoretical framework of co-opetition, according to which both C&C are needed in inter-organizational relationships to allow firms to obtain reciprocal advantages (Bengtsson and Kock, 2000), it could be useful to understand whether (and how) co-opetition could be applied in capacity sharing case. There are many areas of management in which co-opetition may be used. However it is far from being clear in which circumstances a strategy of co-opetition should be used and in which it should not, what the consequences of cooperating and competing at the same time are for all the stakeholders, which variables and parameters involved in co-opetition decisions drive the main results and how managers should design and handle them to make co-opetition work out well. Literature is still far from providing complete analysis of the suitability of co-opetition strategies in given business settings. Research on the topic has just begun. Therefore, a substantial research effort is required to make the picture clearer and provide useful frameworks and tools for firms to decide, design and manage co-opetition. Our goal is to provide a first, and of course not exhaustive, analysis of co-opetition in given production areas. We analyze simultaneous cooperating and competing strategies in capacity sharing case: co-opetition in production, to the best of our knowledge, is not treated at all. The lack of (also) analytical works in this field (see the following section 5) opens up a high number of directions and questions to investigate. Many authors, in fact, have addressed the problem of capacity sharing in multi-unit factory geographically distributed, but nobody using co-opetitive approach. Tonshoff et al. (2000), for example, described a conceptual framework in a decentralized production environment concerning several sales unit and production sites geographically distributed. The capacity allocation plan in this case is made by a mediator in a centralized approach. Ip et al. (2000) proposed an approach concerning planning and scheduling problems in a multi-products manufacturing environment by using Genetic Algorithm. Therefore, the approach is centralized and not applicable in an environment with independent factories. Christie and Wu (2002) proposed an approach to manage the capacity planning in a multi-fabs environment. Each fab is modelled as a single resource with variable production level. Several discrete scenarios are considered in a multi-period, multistage and stochastic programming model. The goal is to minimize the expected mismatch between planned and actual capacity allocation as defined in the scenarios. Chen et al. (2007) proposed model enables a collaborative integration for resource and demand sharing. A negotiation algorithm is utilized to sharing capacity from factory that selling to factory that requires extra capacity. Each factory applies an economic resource planning model based on Genetic Algorithm to improve its local objectives. The model is tested only between two factories: one seller and one buyer. Renna and Argoneto (2008) proposed a distributed approach, for a network of independent enterprises, to facilitate the resources sharing process. The distributed architecture is based on Multi Agent Architecture paradigm and the coordination mechanism was performed by a negotiation protocol. In this work just two performance indexes has been considered: the total profit of the network and the total unsatisfied capacity. Anyway, co-opetition concepts from procurement, marketing and R&D can be also adapted to this area and, therefore, open many spaces for research. To use a sentence from Brandenburger and Nalebuff (1996), in any case, the goal of a firm is to do well for its self-interest, but, especially in a globalized context, to pursue an objective, such as new product development, procurement cost reduction as well as new market entrance, an ally might be necessary. Sometimes, factors such as the knowledge, expertise, technologies, processes and products, the market structure, features and dynamics induce a firm to find the best ally in a competitor getting the way for a win-win result. On the other hand, many other times, a firm can succeed only if able to win out the competition. Therefore a deep analysis of potential benefits and disadvantages of co-opetition strategy is necessary for managers to make proper decisions, design and manage such a new form of business interaction. The doctrine on co-opetition has overcome the positions according to which competition has only negative externalities, while collaboration has exclusively positive effects in an inter-organizational network. In our work, specifically, co-opetition is meant to be jointed in resources procurement among competing firms. That is not the only form of co-opetition in procurement, but it is one of the most common in reality and allows us to easily model the co-opetitive features. In our analysis we investigate advantages and disadvantages of cooperation in presence of competition, by analyzing effects on network, stakeholders and, obviously, on firms. Our results provide useful insights and guidelines for managers involved in this kind of issues, in presence of competition. Because of the novelty of the proposed approach, this work provides basic models to be used as a starting point for further needed developments in the area of production network co-opetition as well as for applications to other business areas.

CO-OPETITION, A LITERATURE OVERVIEW

In spite of the success of the book from Brandenburger and Nalebuff (1996) and although, in the last years, the word co-opetition has been one of the most used and abused in business environments, only few works on the topic are present in the scientific literature. As stated in Dagnino and Padula (2002) scientific investigation on the issue of co-opetition has not gone much farther beyond naming, claiming or evoking it. Some relevant conceptual and/or empirical works in the strategic management literature indirectly face the issue of cooperation among competing firms analyzing strategic alliances formation and firms' networks. One of the main questions addressed by researchers has concerned the reasons behind strategic alliances formation. We have already summarized the most relevant ones. In addition, the Resource Based View perspective contributes to the concept of co-opetition by pointing out that, coming the competitive advantage from resources and capabilities, firms eventually join competitors and interact in networks in order to have access to external critical resources and capabilities and organizational skills (Gulati et al., 2000). The other main question has concerned the keys for an alliance being successful and investigation of failure causes. As pointed out by Bengtsson and Kock (2000), in such stream of literature primarily the cooperative dimension of the relationship is emphasized. In contrast, they argue that both cooperation and competition are simultaneously needed in horizontal relationships since the different relationships may provide the firm with different advantages. In one of the first papers explicitly facing the co-opetition issue, Bengtsson and Kock (2000) use an explorative case study of three industries to show that it is of crucial importance to separate the cooperative part from the competitive part of the relationship. They also show how to separate and manage these two conflicting logics of interaction. More recent contributions organize in a framework different kinds of co-opetitions (Garraffo, 2002), define a typology co-opetition and show how such a strategy may contribute to the value creation or construct a framework to measure intensity and diversity of co-opetition in order to provide some guidelines for building up co-opetitive relationships (Luo, 2007). The papers presented above, even those which explicitly study co-opetition issues, focus on analyzing inter-firms relationships within co-opetition frameworks. None follows the approach suggested by Brandenburger and Nalebuff (1996). They argue for the effectiveness of game theory as a tool to analyze co-opetition, but in spite of their suggestion, only restricted research areas have used game theory to model situations in which both competitive and cooperative behaviors arise. One of these is, surely, the stream of industrial organization which focuses on R&D cooperation in oligopoly: D'Aspremont and Jacquemin (D'Aspremont and Jacquemin. 1988), in their work consider the possibility of cooperation in certain activities among competing firms; specifically cooperation concerns R&D

activities. They compare three cases in presence of spillovers effects: total competition, cooperation in R&D and competition in the final market, total cooperation (collusion), showing that co-opetition and full cooperation do not always have negative impact on social welfare. In the same vein, cooperation in R&D among oligopolists and effects of such cooperation are analyzed in more recent papers (Suetens, 2005). If models of co-opetition are consolidated in the stream of industrial organization that focuses on R&D issues, game-theoretic approaches analyzing co-opetition strategies in other research areas are scattered and their use is far from being considered structured and consolidated. In marketing literature and in agricultural economics, there are some examples of game theoretic approaches, for instance, in analyzing effects and results of joint advertising or co-promotion in presence of competition among partners (Krishnamurthy, 2000; Bass et al., 2005; Isariyawongse et al., 2007). On the other hand, no work in operations and supply chain management seems to focus on co-opetition issues, except some papers (Gurnani et al. 2007) considering simultaneously common and conflicting interests among suppliers and buyers. However, even though in line with the broad definition of co-opetition proposed by Brandenburger and Nalebuff (1996), in these papers, co-opetition does not occur among real competitors, therefore the challenge of analyzing cooperation and competition simultaneously is still missing in operations and supply chain management and in other research areas as well.

This research overcomes the previous researches in the following issues:

- the co-opetion paradigm is efficiently applied to a capacity sharing issue in a network of indipendent plants geographically dispersed;
- a Multi Agent Architecture is properly developed in order to support the network infrastructure;
- game theory is utilised as a policy mechanism to coordinate the network;
- a negotiation approach is developed as a benchmark to compare game theoretical approach;
- a proper discrete event simulation environment, based on open source tool, is developed in order to test the proposed approaches in a dynamic environment.

The conducted experiments allow to highlight the added value of electronic co-opetitive network.

NETWORK GOVERNANCE PROPOSED

Relationships are the focus of substantial investments in time, money and effort and are the means by which knowledge as well as other strategically important resources are both accessed and created. Furthermore, relations are connected to other relations resulting in systems of interdependent relations, henceforth referred to as business networks. Coordination between as well as within relationships becomes a central managerial concern and the means by which the joint productivity of the value system comprising a network of connected business relationships is improved. However, when firms are be involved in co-opetition, a ‘‘coordinator system’’ (or ‘‘intermediate actor’’) is needed in order to plan, coordinate, monitor and appraise inter-organizational relationships. The ‘‘coordinator system’’ should use both formal mechanisms (such as an incentive system) and informal mechanisms (for instance, promoting inter-firm social events for communication and interactions) to adequately balance cooperation and competition. Therefore, the doctrine on competitive relationships has argued that C&C may be managed either through a strategy of separation between competitive and cooperative individuals/subunits or by establishing a ‘‘coordinator system’’ that combines these stimuli in order to maximize system gains from inter-firm co-opetition. Moreover, the emerging new organizational form, allowing for more co-ordination among quasi-independent actors, and, at the same time, more flexibility and autonomy in planning, production and distribution, needs to be properly managed applying right technologies. In the eyes of the authors, the most appropriate tools to model this kind of situation is given by Multi-Agent System (MAS) approach. Being the problem domain particularly complex, large and unpredictable, the only way it can reasonably be addressed is to develop a number of functionally specific and (nearly) modular components (agents) representing firms. This decomposition allows each agent to use the most appropriate paradigm for solving its particular problem. When interdependent problems arise, the agents in the network must coordinate with each other to ensure that interdependencies are properly managed. A MAS can be defined as a loosely coupled network of problem solvers that interact to solve problems that are beyond the individual capabilities or knowledge of each problem solver (Durfee and Lesser, 1989). The characteristics of MAS, and why we use it, are that:

- each agent has incomplete information or capabilities for solving the problem and, thus, has a limited viewpoint;
- there is no system global control;
- data are decentralized; and
- computation is asynchronous.

Sophisticated individual agent reasoning can increase MAS coherence because each individual agent can reason about non-local effects of local actions, from expectations of the behavior of others, or explain and possibly repair conflicts and harmful interactions. Indeed, as previously discussed, a distinguishing feature of a MAS is the fact that the decision making of the agents can be distributed. This means that there is no central controlling agent that decides what each agent must do at each time step, but each agent is to a certain extent responsible for its own decisions. The main advantages of such a decentralized approach over a centralized one are efficiency, due to the asynchronous computation, and robustness, in the sense that the functionality of the whole system does not rely on a single agent. In order for the agents to be able to take their actions in a distributed fashion, appropriate coordination mechanisms must be additionally developed. A typical situation where coordination is needed is among co-opetitive agents that form a network, and through this network they make joint plans, compete for resources and try to pursue common goals. Planning for a single agent is a process of constructing a sequence of actions considering only goals, capabilities, and environmental constraints. However, planning in a MAS environment also considers the constraints that the other agents' activities place on an agent's choice of actions and the constraints that an agent's commitments to others place on its own choice of actions. The development of protocols that are stable (non-manipulable) and individually rational for the agents is the subject of mechanism design. The approaches utilized for this last issue, in this chapter, are negotiation and cooperative game theory. Negotiation is seen as a method for coordination and conflict resolution. It has also been used as a metaphor for communication of plan changes, task allocation, or centralized resolution of constraint violations. The main characteristic of negotiation is the presence of some sort of conflict that must be resolved in a decentralized manner by self-interested agents under conditions of bounded rationality and incomplete information. When one does build negotiation and enforcement procedures explicitly into the model the results depend very strongly on the precise form of the procedures, on the order of making offers and counter-offers and so on. But problems of negotiation are usually more amorphous; it is difficult to pin down just what the procedures are. More fundamentally, there is a feeling that procedures are not really all that relevant; that it is the possibilities for coalition forming. Instead, cooperative game theory abstracts away altogether from procedures and concentrates on the possibilities for agreement. It deals with situations, where a group of players cooperate by coordinating their actions to obtain a joint profit. It is usually assumed that binding agreements between the players are the mean of the cooperation. In the following, we use both this coordination mechanisms to appreciate their weaknesses and strengths in a network co-opetitive circumstances.

THE AGENT-BASED NETWORK MODEL

Sophisticated individual agent reasoning can increase MAS network coherence because each individual agent can reason about non-local effects of local actions, form expectations of the behaviour of others, or explain and possibly repair conflicts and harmful interactions. Numerous works in Artificial Intelligence (hereafter AI) research try to formalize a logical axiomatization for rational agents. This axiomatization is accomplished by formalizing a model for agent's behaviour in terms of beliefs, desires, goals, and so on. An agent will be considered *rational* if it always selects an action that optimizes an appropriate performance measure, given what the agent knows so far. The performance measure is typically defined by the user (the designer of the agent) and reflects what the user expects from the agent in a specific task. For the purpose of this chapter, it will be considered a discrete set of time steps $t = 1,\ldots,n$, in each of which the agent must choose an action a_t from a finite set of available actions A. The complete history of action up to time t to an optimal action is called the *policy* of the agent. Obviously, when the agent takes an action, the "external world" changes as a result of this action. A transition model specifies how the world changes when an action is executed. If the current world state is s_t and the agent takes action a_t, the transition model maps a state-action pair $(s_t; a_t)$ to a single new state s_{t+1}. In classical AI, a goal for a particular task is a desired state of the world. Generally, an agent may hold preferences between any world states. A way to formalize the notion of state preferences is by assigning to each state s a real number $U(s)$ that is called the utility of state s for that particular agent. Formally, for two states s and s' holds $U(s) > U(s')$ if and only if the agent prefers state s to state s', and $U(s) = U(s')$ if and only if the agent is indifferent between s and s'. Intuitively, the utility of a state expresses the *desirability* of that state for the particular agent; the larger the utility of the state, the better the state is for that agent. Equipped with utilities, the question now is how an agent can efficiently use them for its decision making. Namely, what we need is a mechanism able to give the better, in terms of related utility function, *world state* for each involved agent. Specifically two coordination mechanisms are utilised in this chapter: the first is a negotiational one and the second is given by cooperative game theory. They not only differ for the theoretical approach, but for the exchanged information too. The common variables considered for the network formalization are: the market price of the *kth* product, the associated productive cost, related to the price by a mark-up strategy, the productive capacity of each plant and the quantity of each product required by the market. Specifically:

$price_p^k$ is the market price of *kth* product, for the generic *pth* plant;

$\cos t_p^k$ is the production cost of *kth* product, for the generic *pth* plant. It is a function of productive and managerial costs. It also takes into account the efficiency of the plant and its relative geographical dispersion. It is obtained with a mark-up strategy;

C_p^k is the productive capacity of *kth* product, for the generic *pth* plant;

R_p^k is the quantity of *kth* product required by the market, for the generic *pth* plant.

For both approaches, agents are identified and classified in overloaded $OG = \{1,\ldots,i\ldots,N\}$ and underloaded $UG = \{1,\ldots,j\ldots,M\}$. Afterwards each of them respectively compute the capacity it needs to produce a given product *k*, RC_i^k, or the one it can offer, OC_j^k, to other plants of the network. The only variable all agents take into account is the price to pay to obtain -or make over- their capacity.

COORDINATION MECHANISM

Negotiation is one of the most common approaches used as a coordination mechanism when the actors involved have conflicting goals. In particular, when the system is distributed the negotiation is the prevalent approach used. For all these reasons, it is an opportune benchmark to evaluate different coordination mechanism, as the game theoretic one proposed in this chapter. The motivation in finding a new approach lies in the following drawbacks that affect the performance of the negotiation process:

- **Function Definition:** The agents' strategies and the generative function typology for each role (creative or reactive offer) have to be defined and the output strongly depends on this choice;
- **Maximum Number of Rounds:** the performance strongly depends on the number of rounds;
- **Information Exchange:** For example, one agent can simple refuse the proposal or could indicate the terms to improve and so on;
- the negotiation ending criteria, that has to be defined.

Instead, the proposed game theory approach could be implemented in only one step and no strategies have to be designed (a cooperative approach is considered). Therefore, the main advantages are the following:

- the reduction of time to reach an agreement;
- the reduction of information exchange;
- the "intelligence" of the agents can be limited.

Negotiation

The agent architecture in here adopted is the following: each plant belonging to *OG* is represented by a *Capacity Offering Agent* (*COA*) who is in charge of negotiating the capacity with *RCAs* (i.e. *Requiring Capacity Agent*) representing the plants that require it, belonging to the *UG* set. Finally, there is a *Mediator Agent* (*MA*) who is in charge of allowing communication and coordination among *COA*s and *RCAs*.

The negotiation process is characterized by the following constraints (Negotiation constraints):

- the negotiation is a multi-lateral one and it involves one to many agents;
- the negotiation is an iterative process with a maximum number of rounds, r_{max} ; after that an agreement is reached or the negotiation fails;
- during each round (r) the OCA can submit a new counter-proposal (N) to the RCA while, at $r = r_{max}$, it can only accept (A), reject (R) or ask for last counter proposal. Obviously, the RCA answer at generic round r, can be referred as $(A \vee R \vee N)_r$;
- the agreement is reached only if the RCA accepts the OCA counter proposal at round $r < r_{max}$; in this case the agents sign an electronic contract; if there are multiple agreements, the first OCA that satisfies the RCA signs the agreement.

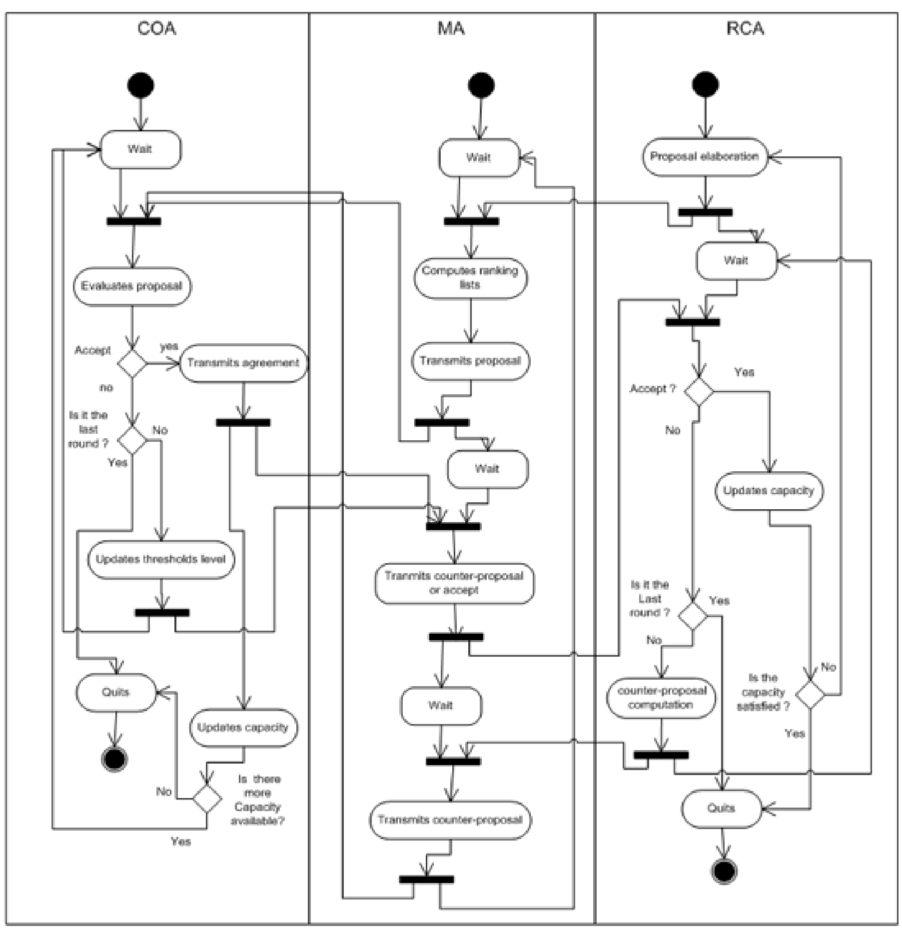
- the agents behavior is assumed to be rationale according to their utility functions;
- the RCA does not know COAs' utility functions and vice versa.

The UML activity diagram of Figure 1 shows the agents' interaction workflow. As the reader can notice three swim lines, corresponding to the above described agents, have been located in the diagram.

Specifically, for the *COA* the following activities can be highlighted:

- **Wait**: the agent is in its initial state of waiting for a proposal (from *RCA*);
- **Evaluates proposal:** the *COA* evaluates the proposal of the *RCA* in terms of required capacity and offered price. At the first round the *COA* communicates the amount of capacity it is willing to offer (the minimum value between the one requested by the *RCA* and its own unused capacity). Subsequently, the *COA* communicates to the *RCA* if it accepts or refuses the proposed price to exchange the promise amount of capacity. Then the *COA* evaluates the proposal of the *RCA* by a threshold function given by (1):

Figure 1. UML Activity Diagram



$$val_j^{k,r_1} = \left[price_j^k - (price_j^k - \cos t_j^k) \cdot \frac{r_1 - 1}{r_{\max} - 1} \right] \cdot M_{ij}^k \quad (1)$$

being

$$M_{ij}^k = \min(RC_i^k, OC_j^k) \quad (2)$$

Expression (1) computed by the COA is a threshold level. Starting from the market price value, during the negotiation the variable r_1 increments and the threshold level decreases until the value of production costs. In this case the generated profit is null.

At this point, the following expression is checked:

$$val_i^{k,r_1} \geq val_j^{k,r_1} \quad (3)$$

If (3) is verified, the *jth* plant supplies the requested capacity of the *ith* plant: they reach an agreement and each ones can update their available capacity.

- **Updates threshold level**: if the *COA* refuses the price submitted by the *RCA*, it updates the threshold level for the next round of negotiation (increases the value of *r* in expression (1)); if the algorithm reached the last round, the *COA* simply quits the negotiation.
- **Updates capacity**: if the negotiation reaches an agreement, the *COA* updates the capacity it owns. In case no more capacity resources are available it quits, otherwise it goes in its Wait state.

The *RCA* performs the following activities:

- **Proposal elaboration**: the *RCA* elaborates a proposal in terms of price and amount of capacity to acquire, and transmits this information to the *MA*. The submitted price is obtained by the following expression:

$$val_i^{k,r_1} = \left[price_i^k - (price_i^k - \cos t_i^k) \cdot \frac{r_{\max} - r_1}{r_{\max} - 1} \right] \cdot M_{ij}^k \quad (4)$$

Expression (4) computed by the RCA starts with a price equal to production costs: the generated profit is the same obtained when the products are produced by its own plant. During the negotiation the price is increased until the value given by the market price. In this case the generated profit is null.

- **Wait**: the *RCA* waits for counter-proposal by the *COAs*.
- **Counter-proposal computation**: if the *COA* refuses the proposal and the negotiation is still running, the *RCA* computes a new counter-proposal (increases the value of *r* in expression (4)). Otherwise (i.e. it is the last round of negotiation), the process ends with no agreement.
- **Updates capacity**: if the negotiation reaches an agreement, the *RCA* updates their information; if the acquired capacity is exactly the required one it quits, otherwise it computes a new proposal for the residual capacity it needs.

The *MA* performs the coordination activities between *COA* and *RCA*. In particular it:

- **Wait**: the *MA* is in its initial state of waiting for a proposal (from the *RCA*).
- **Computes raking list**: the *MA* computes a ranking list among all the plants that requested capacity. The way it does it is depending on several variables; in this research the ranking is done favoring first plants with high need of capacity, allow-

ing them to better satisfy the customers' requests.

- **Transmits proposal**: the *MA* transmits the proposal computed by *RCA*, at the ranking list of *COAs*.
- **Wait**: the *MA* is in state of waiting for the counter-proposal by all the *COAs*.
- **Transmits counter-proposal**: the *MA* transmits the counter-proposal of the *COA* to the *RCA*.

After having uploaded all the necessary values, the generic *ith* plant that does not reach the entire capacity it needs is again inserted in the ranking list. At this point the negotiation starts again. To avoid a deadlock, the agent that does not reach any agreement at the end of the negotiation process is removed by the ranking list.

Game Theory

In this coordination mechanism, all agents communicate each other the capacities they need (offer) and the price they are willing to pay (accept) for the exchange. Properly, the situation can be formalized like an *assignment game*. The outline is a quadruple $AG^t = (OG^t, UG^t, Val^t, O^t)$, where:

- OG^t and UG^t are the already defined vectors at generic time t;
- Val^t is a vector whose element, $val_j^{k,r}$, represents the monetary value (*val*) that *jth* plant, belonging to UG^t, assigns to its own residual capacity to sell it. This value is calculated by the (5):

$$val_j^{k,t} = \cos t_j^k \cdot M_{ij}^k \tag{5}$$

computed as the value under which the *jth* plant does not have any convenience to exchange capacity.

- O^t is a matrix whose generic element $O_{ij}^{\ t}$ is the price the plant $i \in OG^t$ offers for the capacity offered by each player $j \in UG^t$ using the Formula (6):

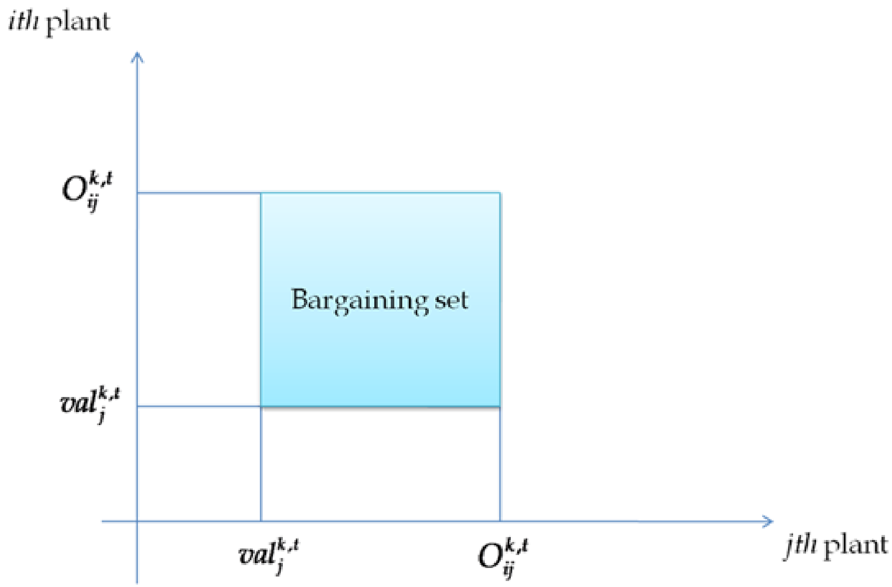
$$O_{ij}^{k,t} = price_i^k \cdot M_{ij}^k \tag{6}$$

computed as the maximum value above which the *ith* plant does not has any convenience to exchange capacity.

Generally speaking, a subset of players is called a coalition. A function v, assigning a value to every possible coalition S, with $v(\emptyset) = 0$, is called a *characteristic function*. The value $v(S)$ is interpreted as the maximum total profit that coalition S can obtain through cooperation. Assuming that the benefit of a coalition S can be transferred between the players of S, the situation is called a cooperative game with transferable utility (TU-game). In reality, the players are not primarily interested in benefits of a coalition but in their individual benefits that they make out of that coalition. A division is a payoff vector specifying for each player its benefit. A division has to be *efficient* and *individually rational*. Efficiency is achieved when some specific criterion is maximized and no allocation of utility could yield a higher value according to that criterion, while individual rationality means that every player gets a quantity of utility at least as much as what he could obtain by staying alone. The set of all individually rational and efficient divisions, the *core game*, is reported in the following Figure 2.

In words, the *core* is the set of imputations under which no coalition has a value greater than the sum of its members' payoffs. Therefore, no coalition has incentive to leave the grand coalition and receive a larger payoff. Specifically for the proposed model, a *transfer utility game* is associated to the quadruple AG^t; in that the whole players are given by $\Gamma^t = OG^t \cup UG^t$ and the characteristic function is defined, given that plants i and j make a coalition, as it follows:

Figure 2. The bargaining set



$$v(i,j) = h_{ij}^{\ t} = \begin{cases} O_{ij}^{k,t} - val_j^{k,t} & \text{if } \ O_{ij}^{k,t} - val_j^{k,t} > 0 \\ 0 & \text{if } \ O_{ij}^{k,t} - val_j^{k,t} \le 0 \end{cases} \tag{7}$$

The set of $h_{ij}^{\ t}$ defines the following assignment problem:

$$\max\left(\sum_{i=1}^{N}\sum_{j=1}^{M} h_{ij}^{\ t}\cdot z_{ij}\right), \tag{8}$$

$$z_{ij} = \begin{cases} 1, \text{if } i \text{ and } j \text{ make a coalition} \\ 0, \text{ otherwise} \end{cases} \tag{9}$$

subject to the following constraints:

$$\sum_{i=1}^{N} z_{ij} = 1; \ \sum_{j=1}^{M} z_{ij} = 1, \text{ if } N = M \tag{10a}$$

$$\sum_{i=1}^{N} z_{ij} = 1; \ \sum_{j=1}^{M} z_{ij} \le 1, \text{ if } N < M \tag{10b}$$

$$\sum_{i=1}^{N} z_{ij} \le 1; \ \sum_{j=1}^{M} z_{ij} = 1, \text{ if } N > M \tag{10c}$$

After having created the assignment among plants, it is necessary to consider the way to subdivide the generated surplus. In order to do this, the optimizing charge belonging to the *core game* is computed. According to Owen's theorem (Owen, 1975), a core allocation, in a linear assignment game, can be obtained starting from an optimal solution of the dual program:

$$\min\left(\sum_{i=1}^{N} y_i + \sum_{j=1}^{M} y_j\right), \tag{11}$$

subject to: $y_i + y_j \ge h_{ij}^{\ t}$ (12)

The residual capacity are then uploaded for each plant involved in the game and the algorithm starts again (t=t+1), formulating a new assignment game, until one of the two following condition -no more workload excess (13) or no more extra capacity (14)- is verified:

$$\sum_{i=1}^{N} RC_i^t = 0 \quad (13)$$

or

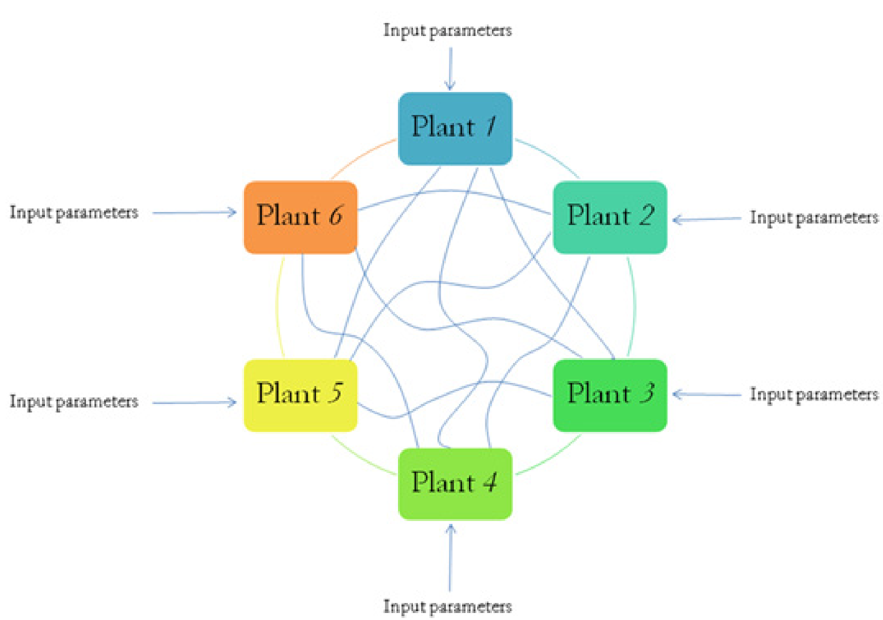
$$\sum_{j=1}^{M} OC_j^t = 0. \quad (14)$$

SIMULATION ENVIRONMENT

It has been developed a distributed simulation environment based on the proposed Multi Agent Architecture to simulate the electronic co-opetitive network represented in Figure 3.

It consists of a simulation environment, developed by using Java development kit package, able to test the functionality of the proposed approaches and to understand the related advantages and/or limits. The modeling formalism in here adopted is a collection of independent agents interacting via messages. This formalism is quite suitable for MAS development. In particular, each object represents an agent and the system evolves through a message sending engine managed by a discrete event scheduler. Specifically, the following objects have been developed: the *COA*, the *RCA*, the *MA* deeply described in the previous section-, the *Scheduler*, the *Model* and the *Statistical agents*. The *Scheduler* agent is in charge of the system evolution by managing the discrete events of the simulation engine. Differently, *Model* agent is in charge of the agents' interaction. Finally, the *Statistical* agent collects the output data -at the end of each run of simulation- and it generates reports and statistical analysis. A proper interface has been developed to connect the *MA* with a Lingo solver in order to solve the mathematical models related to the game theory approach. The simulation network reply 12 periods; in each of them the market price of *kth* product, for the generic *pth* plant $(price_p^k)$ and

Figure 3. The electronic co-opetitive network



the customers demand R_p^k have been randomly generated by using uniform distributions (see Table 2). These parameters are external to the plants and they only depend on the market conditions in which each plant operates. Differently, the "internal" parameters (C_p^k, plant capacity and $\cos t_p^k$, the production cost of *kth* product, for the generic *pth* plant) have been randomly generated at the beginning of each period. Because of the use of a statistical distribution for all parameters, each experiment class has been replicated in order to achieve a confidence degree equal to 95% and a 5% of confidence interval for each performance index (see Table 3). In order to test the proposed coordination strategies, under different network conditions, two levels of variability for each parameter have been considered: *low* (10% of variance) that leads to homogenous network (i.e. plants and markets with similar characteristics) and *high* (50% of variance) that leads to an inhomogeneous network (plants and markets have different characteristic). Thus, combining different levels of all four parameters, 16 simulation classes of experiments have been obtained (see the following Table 1).

The following performance measures have been considered to compare the proposed coordination strategies:

- the *total profit* (**TP**) reached by the whole network; it has been computed as the sum

Table 1. Experimental classes

Experiment classes	Mark-up $\cos t_p^k$	Plant capacity C_p^k	Customer demand R_p^k	Price $price_p^k$
1	L	L	L	L
2	L	L	L	H
3	L	L	H	H
4	L	L	H	L
5	L	H	L	L
6	L	H	L	H
7	L	H	H	H
8	L	H	H	L
9	H	L	L	L
10	H	L	L	H
11	H	L	H	H
12	H	L	H	L
13	H	H	L	L
14	H	H	L	H
15	H	H	H	H
16	H	H	H	L

Table 2. Input parameters (distribution of)

	low	*high*
Mark-up $(\cos t_p^k)$	Uniform[0.9,1)	Uniform[0.7,1)
Plant capacity (C_p^k)	N(100,10%)	N(100,50%)
Customers demand (R_p^k)	N(100,10%)	N(100,50%)
Price $(price_p^k)$	N(10,10%)	N(10,50%)

Table 3. Average results over all experimental classes

Performance	**no network *vs* Negotiation**	**negotiation *vs* game theory**
TP	17.02%	14.89%
TUD	-44.62%	11.29%
TUC	-43.87%	10.56%
PDI	-55.78%	-61.59%
NAL	-	-31.18%
DAL	-	-85.28%
TV	-	47.09%
DOT	-	-94.50%

of each single profit generated by all the plants of the network;

- the *total unsatisfied demand* (**TUD**): it is the difference existing between the quantity of products required by the market and the one the network has been able to satisfy. It could be considered as a customer performance;
- the *total unutilised capacity* (**TUC**): it is the difference existing between the whole capacity of the network and the unallocated capacity;
- the *profit distribution index* (**PDI**): this value computes the capability of the proposed coordination mechanisms to distribute the whole profit among the plants of the network. To compute it, first is necessary to evaluate the average profit of the network as reported in (15):

$$average\ profit = \frac{1}{P}\sum_{P=1}^{P} profit_p \tag{15}$$

being P the number of plants and $profit_p$ the generic profit reached by the *pth* plant. Then, the index is calculated as:

$$PDI = \sum_{p=1}^{P} \frac{\left|profit_p - average\ profit\right|}{average\ profit} \tag{16}$$

- the *number of activated links* (**NAL**) among plants. This performance is an index explaining how much could be complicated the logistic management of the network;
- the *transactions value* (**TV**). It is the quantity of money exchanged among plants, depending on the particular coordination mechanism.
- the distribution of the activated links (**DAL**). This index point out if some preferential links is established. The performance index is computed as it follows. First we compute the number of average links as reported in (17):

$$average\ links = \frac{1}{P \cdot \frac{1}{2}(P-1)} \sum_{p=1}^{P}\sum_{w>p}^{P} link_{wp} \tag{17}$$

being $P \cdot \frac{1}{2}(P-1)$ the total amount of links that can be potentially activated among P plants. The variable $link_{wp}$ is a binary one. It is equal to the unit when the link is activated, otherwise it is null. Then we compute the unbalance index as (18):

$$DAL = \sum_{p=1}^{P} \sum_{w>p}^{P} \frac{\left| link_{wp} - average\, links \right|}{average\, links} \tag{18}$$

The distribution of transactions among plants (**DOT**). This index emphasizes the evaluation of preferential links by considering the amount of exchanged money. The performance index is computed evaluating (19):

$$average\, transactions = \frac{1}{P \cdot \frac{1}{2}(P-1)} \sum_{p=1}^{P} \sum_{w>p}^{P} transactions_{wp} \tag{19}$$

being $transactions_{wp}$ the amount of money exchanged between the *wth* and *pth* plants. Then the index is computed as (20):

$$DOT = \sum_{w=1}^{P} \sum_{w>p}^{P} \frac{\left| transaction_{wp} - average\, transactions \right|}{average\, transactions} \tag{20}$$

RESULTS

All the simulations have been conducted with a network of six plants. Each index –except the mark-up that comes up from an uniform distribution- is considered randomly drawn from a normal distributions *N(μ;σ)* whose parameters are reported in Table 2 (*low* and *high* are referred to the standard deviations size):

The difference among the standard deviations of the statistical input is done to easily highlight the difference between the coordination approaches in both homogenous (*low level*) and inhomogeneous (*high level*) network. The simulation results (the average over the all experimental classes) have been reported in Table 3. Specifically: in the first column the performance indexes are reported using the notation described in the previous paragraph; the second column reports the percentage difference between results obtained in case of no relations among plants and the one in which the network is coordinated by using a negotiation approach. Last column shows the dissimilarity obtained comparing the results coming from negotiation and game theory. The considered performances have been obtained as average values over all the experimental classes.

From the analysis of the previous results, the following considerations can be summarised:

- coordination among plants is a real added value both for plants belonging to the network and for customers. Indeed, comparing the performances with the case in which no relations exist among plants, **TP** increases and, at the same time, **TUC** decreases;
- the negotiation approach leads to a better performances in terms of **TUD** and **TUC**;
- game theory coordination leads to better performance of **TP** reached by the network;
- game theory reduces the number of activated links (**NAL**): that means that this coordination approach simplifies the logistic management among plants involved in the network. At the same time, the high value of TV highlight that, in spite of this reduction, this coordination mechanism increases the amount of each single transaction;
- Considering the unbalanced indexes, **PDI** and **DAL**, as the reader can notice, game theory outperforms the negotiation mechanism: that means that there's a better distribution of the added value of the network and a simpler logistic management.

More relevant results over all the experimental classes are graphically showed in the following Figures 4, 5 and 6.

Figure 4. Profit distribution over 16 experimental classes

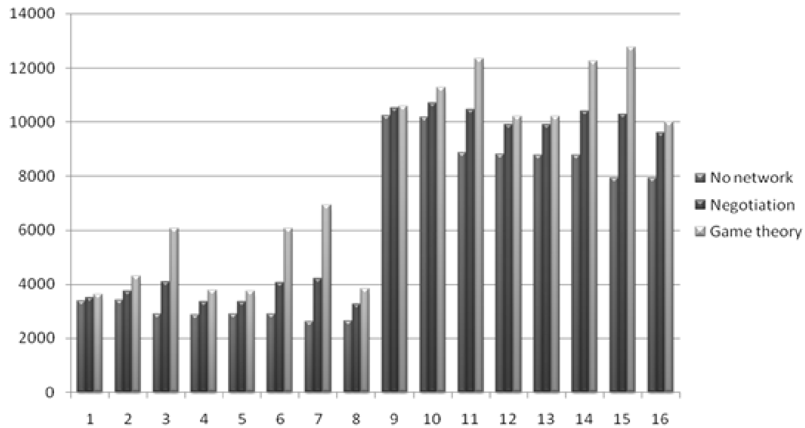


Figure 5. Unsatisfied demand (TUD) over 16 experimental classes

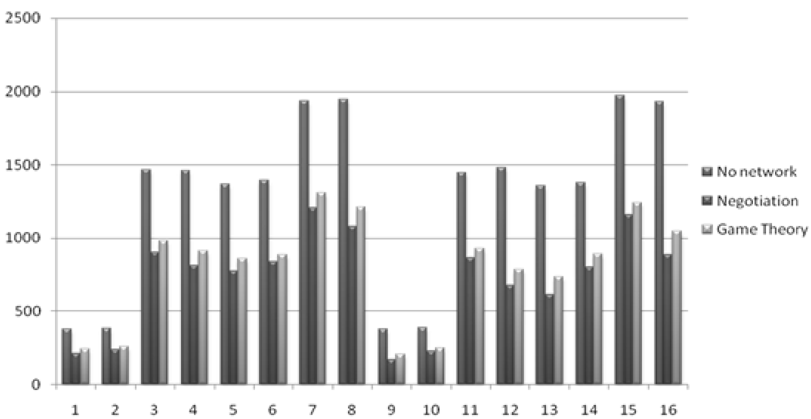


Figure 6. Transactions distribution (DOT) over 16 experimental classes

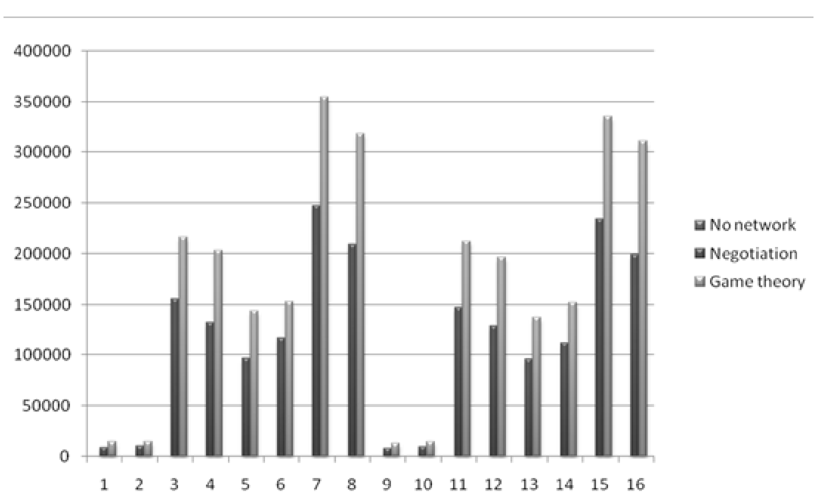


Figure 4 shows the amount of **TP**-reached by the network- for each experiment. Differences among configurations become relevant in the following classes: 3,6,7,11,14 and 15. Being the mark-up variability low (in 3,6 and 7) and high (in 11,14 and 15), it is possible assert that the highlighted differences between negotiation and game theory are not affected by the variability of this parameter. Moreover, they have a comparable reaction when the high level variability of the price is combined with a high variability of the demand (see classes 3 and 11) or of the capacity (see classes 6 and 14), or both the two (see classes 7 and 15).

Figure 5 shows the **TUD** performance of the network, for each experiment class. **TUD** is very low for classes with low variability of input parameters (see classes 1,2,3 and 4). In other cases, the coordination approaches lead to comparable values of this performance, with a little bit minor value for the negotiation approach.

Figure 6 shows the **DOT** performance of the network. For experimental classes with low variance of input parameters (see 1,2, 9 and 10) the amount of transactions is very low. For all the other, the value of transactions obtained using game theory is always greater than in the negotiation case.

SUMMARY, CONCLUSION AND FUTURE DEVELOPMENT

The chapter deals with the real added value in an electronic co-opetitive network of plants that can exchange productive capacity among them. Results of this research can be located at two levels. Concerning the specific plant coordination problem here addressed, the following conclusion can be drawn:

- different coordination strategies lead to a different performances for the networked enterprises; therefore the coordination strategy play a fundamental role for the goal the network would pursue;
- the proposed negotiation mechanism leads to a better performance for the customers: the unsatisfied demand is very low. From the network point of view, the profit is minor than in the game theory case and the distribution among plants is less homogeneous. Moreover, the negotiation approach increases the number of activated links with a minor amount of individual monetary transaction. This characteristics leads to make more difficult the management of the network from the logistic point of view. Finally, the negotiation approach implies the creation of preferential links that could marginalize some plants.
- The proposed game theory approach leads to better performances, from the networked plants point of view. In particular, this approach outperforms the negotiation in terms of generated total profit, uniformity of profit distribution and homogeneity of monetary transactions concerning the activated links. The logistics problem is simpler than in the previous case, because game theory creates minor links with higher individual monetary exchange.

At strategic level, this research shows that:

- Multi Agent System is a suitable approach to implement a distributed architecture for the capacity exchange issue that often characterize electronic co-opetitive network of enterprises;
- Discrete event simulation is a powerful tool to test coordination strategies and highlight the real added value of these approaches in an electronic co-opetitive network. This tool reduces the risk related to the ICT investment and agent based technology. Moreover it can encourage enterprises to

adopt the "stay together" approaches for real business applications.

Obviously the network is not a permanent fixture, but is something that is either being formed or might change. From this point of view, and generally speaking, the allocation of value at a given network can and should depend on the value that might accrue to alternative potential networks. In particular, evaluating the contribution to value of a given link depends on the contribution of that link to various networks, and perhaps more importantly, which to what extent other links might serve as a substitute. To understand why this issue arises it is important to recognize that larger networks might have higher costs associated with them than some smaller networks. For instance, the value generated by the complete network might be much less than the value generated by some sub-network. This means that generally, the most efficient network (in a value maximizing sense) might not be the complete network. This introduces some important considerations in the allocation of value to consider in future development of this work.

REFERENCES

Afuah, A. (2000). How much do your co-opetitors' capabilities matter in the face of technological change? *Strategic Management Journal*, *21*(3), 387–404. doi:10.1002/(SICI)1097-0266(200003)21:3<397::AID-SMJ88>3.0.CO;2-1

Ahuja, G. (2000). The duality of collaboration: inducements and opportunities in the formation of interfirm linkages. *Strategic Management Journal*, *21*(3), 317–343. doi:10.1002/(SICI)1097-0266(200003)21:3<317::AID-SMJ90>3.0.CO;2-B

Anand, B., & Khanna, T. (2000). Do firms learn to create value? The case of alliances. *Strategic Management Journal*, *21*(3), 295–315. doi:10.1002/(SICI)1097-0266(200003)21:3<295::AID-SMJ91>3.0.CO;2-O

Ardichvili, A., Cardozo, R., & Ray, S. (2003). A theory of entrepreneurial opportunity identification and development. *Journal of Business Venturing*, *18*(1), 105–123. doi:10.1016/S0883-9026(01)00068-4

Bass, F. M., Krishnamoorthy, A., Prasad, A., & Sethi, S. P. (2005). Generic and Brand Advertising Strategies in a Dynamic Duopoly. *Marketing Science*, *24*(4), 556–568. doi:10.1287/mksc.1050.0119

Baum, J., Calabrese, T., & Silverman, B. (2000). Don't go it alone: alliance network composition and startups' performance in Canadian biotechnology. *Strategic Management Journal*, *21*(3), 267–294. doi:10.1002/(SICI)1097-0266(200003)21:3<267::AID-SMJ89>3.0.CO;2-8

Bengtsson, M., & Kock, S. (1999). Cooperation and competition in relationship between competitors in business network. *Journal of Business and Industrial Marketing*, *14*(3), 178–193. doi:10.1108/08858629910272184

Bengtsson, M., & Kock, S. (2000). Coopetition in business networks to cooperate and compete simultaneously. *Industrial Marketing Management*, *29*(5), 411–426. doi:10.1016/S0019-8501(99)00067-X

Boari, C., & Lipparini, A. (2000). Networks within industrial districts: organising knowledge creation and transfer by means of moderate hierarchies. *Journal of Management and Governance*, *99*(3), 339–360.

Brandenburger, A., & Nalebuff, B. (1996). *Co-Opetition: A Revolution Mindset That Combines Competition and Cooperation*. New York: Harper Collins Business.

Brandenburger, A. M., & Nalebuff, B. J. (1996). *Co-opetition*. New York: Doubleday Dell Publishing Group.

Chen, J. C., Wang, K. J., Wang, S. M., & Jang, S. J. (2008). Price negotiation for capacity sharing in two-factory environment using genetic algorithm. *International Journal of Production Research*, *46*(7), 1847–1868. doi:10.1080/00207540601008440

Christie, R. M. E., & Wu, S. D. (2002). Semiconductor capacity planning: Stochastic modeling and computational studies. *IIE Transactions*, *34*(2), 131–144. doi:10.1080/07408170208928856

Cooper, A. (2002). Networks, alliances and entrepreneurship. In Hitt, M., Ireland, D., Camp, M., Sexton, D. (Eds.), *Strategic Entrepreneurship,203-222* Oxford, UK: Blackwell, Oxford.

D'Aspremont, C., & Jacquemin, A. (1988). Cooperative and Noncooperative R&D in Duopoly with Spillovers. *The American Economic Review*, *78*, 1133–1137.

Dagnino, G. B., & Padula, G. (2002). Coopetition Strategy. A New Kind of Interfirm Dynamics for Value Creation. *EURAM Second Annual Conference – "Innovative research in Management"*, Stockholm, 9-11 May.

Dollinger, M. (1985). Environmental contacts and financial performance of the small firm. *Journal of Small Business Management*, *23*(1), 24–30.

Doz, Y., Olk, P., & Ring, P. (2000). Formation processes of R&D consortia: which path to take? Where does it lead? *Strategic Management Journal*, *21*(3), 239–266. doi:10.1002/(SICI)1097-0266(200003)21:3<239::AID-SMJ97>3.0.CO;2-K

Durfee, E. H., & Lesser, V. (1989). Negotiating task decomposition and allocation using partial global planning. In Gasser, L., & Huhns, M. (Eds.), *Distributed artificial intelligence* (*Vol. 2*, pp. 229–244). San Francisco, CA: Morgan Kaufmann.

Freel, M. (2000). External linkages and product innovation in small manufacturing firms. *Entrepreneurship and Regional Development*, *12*, 245–266. doi:10.1080/089856200413482

Galaskiewicz, J., & Zaheer, A. (1999). Networks of competitive advantage. In Andrews, S., & Knoke, D. (Eds.), *Research in the Sociology of Organizations* (pp. 237–261). Greenwich, CT: JAI Press.

Garraffo, F. (2002). Types of Coopetition to Manage Emerging Technologies. *EURAM Second Annual Conference – "Innovative research in Management"*, Stockholm, 9-11 May.

Granovetter, M. (1985). Economic action and social structure: A theory of embeddedness. *American Journal of Sociology*, *91*, 481–510. doi:10.1086/228311

Gulati, R. (1998). Alliances and networks. *Strategic Management Journal*, *19*(4), 293–317. doi:10.1002/(SICI)1097-0266(199804)19:4<293::AID-SMJ982>3.0.CO;2-M

Gulati, R., Nohria, N., & Zaheer, A. (2000). Strategic networks. *Strategic Management Journal*, *21*(3), 203–215. doi:10.1002/(SICI)1097-0266(200003)21:3<203::AID-SMJ102>3.0.CO;2-K

Gurnani, H., Erkoc, M., & Luo, Y. (2007). Impact of Product Pricing and Timing of Investment Decisions on Supply Chain Co-opetition. *European Journal of Operational Research*, *180*, 228–248. doi:10.1016/j.ejor.2006.02.047

Huxham, C., & Vangen, S. (2005). *Managing to collaborate. Theory and practice of collaborative advantages*. London, New York: Routledge.

Ip, W. H., Li, Y., Man, K. F., & Tang, K. S. (2000). Multi-product planning and scheduling using magnetic algorithm approach. *Computers & Industrial Engineering*, *38*, 283–296. doi:10.1016/S0360-8352(00)00044-9

Isariyawongse, K., Y. Kudo, V. J. Tremblay. (2007). Generic and Brand Advertising in Markets with Product Differentiation. *Journal of Agricultural and Food Industrial Organization* *5*(1), Article 6.

Jarillo, C. (1989). Entrepreneurship and growth: the strategic use of external sources. *Journal of Business Venturing*, *4*, 133–147. doi:10.1016/0883-9026(89)90027-X

Kogut, B. (2000). The network as knowledge: generative rules and the emergence of structure. *Strategic Management Journal*, *21*(3), 405–425. doi:10.1002/(SICI)1097-0266(200003)21:3<405::AID-SMJ103>3.0.CO;2-5

Kotzab, H., & Teller, C. (2003). Value-adding partnerships and coopetition models in the grocery industry. *International Journal of Physical Distribution & Logistics Management*, *23*(3), 268–281. doi:10.1108/09600030310472005

Krishnamurthy, S. (2000). Enlarging the Pie vs Increasing One's Slice: an Analysis of the Relationship between Generic and Brand Advertising. *Marketing Letters*, *11*, 37–38. doi:10.1023/A:1008146709712

Lechner, C., & Dowling, M. (2000). The evolution of industrial districts and regional networks: the case of the biotechnology region Munich/Martinsried. *Journal of Management and Governance*, *99*(3), 309–338.

Lechner, C., & Dowling, M. (2003). Firm networks: external relationships as sources for the growth and competitiveness of entrepreneurial firms. *Entrepreneurship and Regional Development*, *15*, 1–26. doi:10.1080/08985620210159220

Luo, Y. (2004). A coopetition perspective of MNC-host government relations. *Journal of International Management*, *10*(4), 431–451. doi:10.1016/j.intman.2004.08.004

Luo, Y. (2005). Toward coopetition within a multinational enterprise: A perspective from foreign subsidiaries. *Journal of World Business*, *40*(1), 71–90. doi:10.1016/j.jwb.2004.10.006

Luo, Y. (2007). A Coopetition Perspective of Global Competition. *Journal of World Business*, *42*, 129–144. doi:10.1016/j.jwb.2006.08.007

Meder, K. (2005). Case Study SupplyOn: an e-Marketplace from Suppliers for Suppliers. *E-Business W@tch*.

Oliver, A. L. (2004). On the duality of competition and collaboration: Network-based knowledge relations in the biotechnology industry. *Scandinavian Journal of Management*, *20*(1-2), 151–171. doi:10.1016/j.scaman.2004.06.002

Owen, G. (1975). On the core of linear production games. *Mathematical Programming*, *9*, 358–37. doi:10.1007/BF01681356

Quintana-Garcia, C., & Benavides-Velasco, C. A. (2004). Cooperation, Competition, and Innovative Capability: a Panel Data of European Dedicated Biotechnology Firms. *Technovation*, *24*, 927–938. doi:10.1016/S0166-4972(03)00060-9

Renna, P., & Argoneto, P. (2008). *Capacity Allocation in Multi-Site Factory Environment: a Multi Agent Systems Approach. Intelligent Computation in Manufacturing Engineering Innovative and Cognitive Production Technology and Systems, 23 - 25 July*. Italy: Naples.

Rowley, T., Behrens, D., & Krackhardt, D. (2000). Redundant governance structures: an analysis of structural and relational embeddedness in the steel and semiconductor industries. *Strategic Management Journal*, *21*(3), 369–386. doi:10.1002/(SICI)1097-0266(200003)21:3<369::AID-SMJ93>3.0.CO;2-M

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, *25*, 217–226. doi:10.2307/259271

Suetens, S. (2005). Cooperative and Noncooperative R&D in Experimental Duopoly Markets. *International Journal of Industrial Organization*, *18*, 63–82. doi:10.1016/j.ijindorg.2004.11.004

Tonshoff, H.K., Seilonen, L., Teunis, G. and Leitao, P. (2000). A mediator-based approach for decentralized production planning, scheduling and monitoring. *Report of the Espirit program EP 24986.*

ADDITIONAL READING

Belussi, F., & Arcangeli, F. (1998). A typology of networks: flexible and evolutionary firms. *Research Policy*, *27*, 415–428. doi:10.1016/S0048-7333(98)00074-2

Billand, P., & Bravard, C. (2004). Non-cooperative networks in oligopolies. *International Journal of Industrial Organization*, *22*, 593–609. doi:10.1016/j.ijindorg.2004.01.004

Chatterjee, P. (2002). Interfirm alliances in online retailing. *Journal of Business Research*, *57*, 714–723. doi:10.1016/S0148-2963(02)00362-4

Lechner, C., Dowling, M., & Welpe, I. (2006). Firm networks and firm development: The role of the relational mix. *Journal of Business Venturing*, *21*, 514–540. doi:10.1016/j.jbusvent.2005.02.004

Roma, P. (2009). *Models of Co-opetition in group-buying and advertising*. PhD Thesis, University of Palermo, Italy.

This work was previously published in Electronic Supply Network Coordination in Intelligent and Dynamic Environments: Modeling and Implementation, edited by Iraj Mahdavi, Shima Mohebbi and Namjae Cho, pp. 291-318, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 63
Evaluation of Remote Interface Component Alternatives for Teaching Tele-Robotic Operation

Goldstain Ofir
Tel-Aviv University, Israel

Ben-Gal Irad
Tel-Aviv University, Israel

Bukchin Yossi
Tel-Aviv University, Israel

ABSTRACT

The Internet developments as well as the increase in PCs' capabilities and bandwidth capacity have made remote learning through the internet a convenient and practical learning preference, leading to a variety of new learning interfaces and methods.

This chapter discusses a remote learning study conducted at the Computer-Integrated-Manufacturing (CIM) Laboratory in Tel-Aviv University. The goal is to provide remote end-users with an interface that enables them to teleoperate a robotic arm in conditions as close as possible to hands-on operation in the laboratory. This study evaluates the contribution of different interface components to the overall performance and the learning ability of potential end-users. Based on predefined experimental tasks, the study compares alternative interface designs for teleoperation. The three performance measures of the robot operation task are (1) the number of steps that are required to complete the given task, (2) the number of errors during the execution stage, and (3) the improvement rate of users. Guidelines for a better design of remote learning interfaces in robotics are provided based on the experimental results.

DOI: 10.4018/978-1-4666-1945-6.ch063

INTRODUCTION

This chapter focuses on the design of an interface for remote learning of robotics operations. The interface design, which is supported by technical guidelines, is general and applicable for a wide variety tools for teaching tele-robotic operation. It differs from previous research in the field, which often focuses on a specific applicative interface.

The proposed interface includes aspects of remote manipulation of robots with aspects of remote learning. The motivation for such integration is to enable users to practice not only the remote activation of a robotic cell but also the availability of learning, redesigning and optimizing the work plan in the cell. The chapter starts by considering three possible design schemes: a "Home-based," a "Lab-based" and a "Website-based." It identifies different interface components that support a remote telerobotic-learning. Then it measures and evaluates the interactions among these components as well as their effects and usability within a proposed remote learning interface. Such an evaluation is conducted by running a set of experiments, requiring the users to execute specific robotic tasks from a remote location while examining their performance over various interface settings. The performance of the remote users is also compared with hands-on operation, which is used as a benchmark setting.

The evaluation tool of the web-based interface for the telerobotic learning is called the *Test-Oriented-Interface* (TOI). As the chapter unfolds, elements within this interface are evaluated, focusing on their contribution to the remote learning assignments. A full set of guidelines for designing a remote learning interface is extracted from the evaluation of the TOI. The objective of these guidelines is to maximize the benefits obtained from the interface for the users (e.g., students) as well as for the hosting institute (e.g., university). Finally, we present how the new web-interface for remote learning of robotic operations is implemented and fully operated in the CIM laboratory.

BACKGROUND

Remote Learning Interfaces for Robotics

Remote control and manipulation of robots has been used to perform predetermined tasks, often in a hostile, unsafe, inaccessible or remote environment (Siegwart and Saucy, 1999 and Bukchin et al., 2002). NASA, for example, keeps track and provides free access to active telerobotic systems through the NASA Telerobotics Web-page.

Architecture of a WWW-based system for a remote telerobotic operation was presented in 1999 by Belousov et al. (1999). Their system was mainly oriented for reliability and efficiency and was based on a 3D Java visualization tool that overcame bandwidth restrictions that existed at the time. Belousov et al. (2001) presented a similar architecture with an addition of a tool that supports the remote programming of the robot.

Among many variations of teleoperation systems, one can mention Wang and Liu (2004) with a teleoperation paradigm for Human-Robot interaction, Hu et al. 2001)) with a system for remote controlling a robot with visual feedbacks over a simulated map, Kofman et al. (2005) with a hand-arm-gesture method for teleoperation, Hu et al. (2001) with a pioneering work networked telerobotic systems for tele-training, and Siegwart and Saucy (1999) with a modular framework for mobile robots on the web and many other applications.

New integration protocols were used to combine 3D simulation tools with remote control and manipulation interfaces, enabling the management of complicated tasks in flexible robotic cells. Candelas et al. (2005) present a system focused on the training of kinematics and trajectory design of robotic arms. Their work was among the first to use a learning platform with full interactivity in the teleoperation process. Michau et al. (2001) present in detail the expected benefits of web-based learning for engineers. They express the need for remote learning tools within virtual laboratories, stating that although simulations cannot replace real experiments, remote laboratories provide new ways for practicing hands-on-experiments (Gravier et al., 2008).

Integrating simulations within real implementation activities is considered a necessity in contemporary engineering education (Khan and Al-Kahtani, 2002; Fernandes and Martins, 2001). An example for such integration is found in Calkin et al. (1998), with visualization, simulation and control of a robotic system relying on internet technology. This virtual learning mechanism was later referred to by Goldstain et al. (2007) as the "Home-Based" design scheme. Similarly, Puente et al. (2000) use simulations as a learning tool when suggesting general system architecture. Yang et al. (2004) introduce an internet-based teleoperation scheme of a robot manipulator for educational purpose. Their system integrates a virtual off-line simulation with an actual teleoperation module, including a visual feedback. In their conclusion they suggest the development of a more general control system that later was presented in Goldstain et al. (2007) as the "Web-based" design scheme.

Enrique Sucar et al. (2005) consider virtual modules as a primary tool in teaching robotics. In their work, a virtual laboratory based on simulation was developed and assessed without evaluating the required interface components as done in this chapter. A modern, fully developed interface for remote learning and programming of a robot arm was also presented by Marin et al. (2005).

Remote Interfaces Design

Siegwart and Saucy (1999) describe the basic specifications and major difficulties when designing an interactive platform for remote learning of robotics systems. Their suggested modules include a video feedback module, a robot guidance module, and a virtual representation module.

Enabling a user to learn and optimize a work plan in addition to remotely operating given robotic tasks requires more than basic manipulation tools for remote control (Kahn et al., 2002). A three-dimensional (3D) simulation tool is one of the most popular tools when dealing with "on-site" learning (Candelas et al., 2005; Enrique Sucar et al., 2005). Candelas et al.(2005), Tzafestas et al. (2006), Marin et al. (2005), and others offered different variations of both off-line and online 3D simulation tools. More advanced simulation tools, like the one used in Goldstain et al. (2007), provide another important feature for the learning process, which is the ability to create and record a program for the simulated system and then implement it to run the physical system itself.

The main feedback for a remote operation of a robotic cell is often considered to be visual feedback (Hu et al., 2001; Eliav et al., 2005; Aktan et al., 1996). Unlike local settings, the remote setting often relies on various visual sensors to provide visual feedback (Bukchin et al., 2002; Bochiccio and Longo, 2009). In virtual laboratories in particular, this feedback is gained through a 3D model, as implemented in Belousov et al. (1999) and Belousov et al. (2001) using Java. Another, more advanced feedback for robotics is presented in Tzafestas et al. (2006) and Goldstain et al. (2007). These works consider a positional feedback, providing the user with valuable information regarding the positioning of each axis in the robotic arm. This type of feedback can be used to reconstruct the robot's movement or even to completely re-evaluate the robotic cell layout.

Adams (2002) suggests critical considerations for human-robot interface development. His concepts of User Centered Design and Situation Awareness guided the design of the proposed framework. Goldstain et al. (2007) presented a methodology, which we used for the design of experiments presented in the next section. Their suggested methodology is based on the framework presented by Yang et al. (2004) and by Chen et al.(2001), preaching for the use of virtual laboratories as an essential tool for learning, prior to the execution of "on hand" experiments. This chapter follows the above mentioned guidelines and suggestions. It focuses on the analysis of the main components in an interface for remote-learning of robotic operations and provides guidelines for an efficient design of such tool.

Integrating Remote-Learning and Teleoperation

The main challenge addressed in this chapter is how to enable multiple remote users (e.g. students) to simultaneously design and optimize the operation of a robotic cell. The users are required to learn, mainly through a "trial-and-error" process, the appropriate way to operate the robotic cell previous to its online activation. This supports cognitive (or mental) learning, as opposed to a motoric (or mechanical) learning, which is required when operating a robotic cell on site (Bukchin et al., 2002; Calkin et al., 1998; Cooper et al., 2002).

There are several alternative approaches that can be considered for remote teaching of robotic operation (Bukchin et al., 2002; Calkin et al., 1998; Tzafestas et al., 2006). The first that we considered is the *Lab-based* approach which uses designated software packages, i.e. Virtual VNC, to enable a full remote control of the robotic cell's controller PC, while receiving all feedbacks remotely. The Lab-based approach allows users to operate and react as if they are physically located in the laboratory. On the other hand, it requires a considerably broad bandwidth from both the client's and the server's computers. It also lacks the possibility of administrating several users simultaneously, and bares with it many security IT problems.

Second to be considered is a *Home-based* approach. This approach provides the users with a simulation program and an accurate scaled 3D model file of a cell's layout, enabling them to initiate the learning at home without being dependent on the robot's availability at the lab. In this approach the users are required to simulate and optimize the entire robotic operations on their home computer and then send the final programmed tasks to the robot's controller PC to be executed while they can observe the execution through a video feed (Calkin et al., 1998). The Home-Based approach, although safer and cheaper in terms of bandwidth demand, lacks the interactivity part that is so essential to support a real "trial-and-error" learning process.

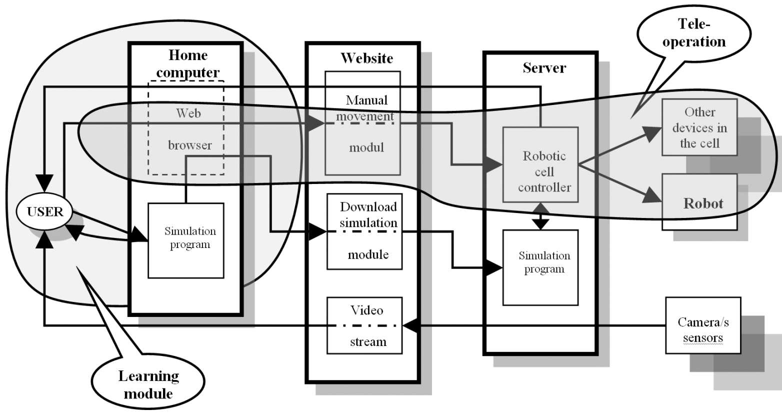
Finally, a third approach is investigated by combining the above-mentioned two approaches. It is named the *Web-based* approach and relies on a specially-designed website that is integrated with a 3D simulation program installed on the user's computer. Goldstain et al (2007) claim that this approach has several significant advantages: it is safer, it enables future developments for other types of lab equipment, it supports administrating multiple users, it controls the system's usage, and it keeps the laboratory's computer invisible to the end users, thus, reducing the associated safety problems (Candelas et al., 2005; Chiculita and Frangu, 2002). The methodology used in the web-based design relies on client-server interactions (Chen et al., 1999; Chiculita and Frangu, 2002; Saliah et al., 2000). Such interactivity was found to be essential to provide the users with the necessary information to refine their actions to achieve a more accurate performance of the required task. A survey conducted among 60 users of a beta version of this website and presented in Goldstain et al. (2007) indicated a high level of user satisfaction.

METHODOLOGY

In this section a conceptual methodology is proposed to integrate a remote learning process with teleoperation, based on the web-based approach presented above. The integration of remote learning and teleoperation within the web-based approach is presented in Figure 1.

The web-based solution includes the following modules: (1) a user-client PC running a simulation

Figure 1. "Web-based" design scheme



program and a web browser; (2) a website containing video capabilities, manual movement controls and files-upload protocol; and (3) a web server running the robotic cell controller as well as the simulation program. These combined modules support both the learning through "trial-and-error" and a teleoperation mechanism.

Basic Modules of Remote Learning of Robotic Operations

Enabling a user to learn, operate and optimize given robotic tasks requires more than basic manipulation interface for remote control. For example, a 3D simulation tool is one of the most popular tools for dealing with hands-on learning (Candelas et al., 2005; Enrique Sucar et al., 2005). The simulation provides the users with the ability to analyze their performance and to redesign the process up to the required level. Advanced simulation tools provide another important feature to the learning process–namely, the ability to create and record a robotic program and then apply it to actually run the physical system. When using a recorded program, the users can easily implement changes to the system layout or to the sequence of performed actions by using historical knowledge. Both the simulation and the optimization processes are based on the interaction between the human operator and the manipulation interface of the robot itself. A feedback from the system allows the users to change and better design their program, thus supporting the process and improving learning skills and abilities.

The basic feedback for a remote operation of a robotic cell is the visual feedback (Eliav et al., 2005; Aktan et al., 1996; Hu et al., 2001). In remote systems, several visual sensors can be used to provide such feedback, often a closed-loop TV system or a streaming video. The latter provides the operator with a somewhat rough knowledge regarding the work cell's layout and the system state following each phase of the operation; however, it is rarely accurate enough to enable fine tuning of the robotic arm tasks, as it provides the users with a two dimensional picture of a three dimensional reality. When considering the visual feedback module, the main objective is to obtain a clear online feed from a few different view angles of the work cell. An issue that remains unsolved is the visual feedback gap between the 3D reality vs. a 2D image (Doulgeri. and Matiakis, 2006; Yang et al., 2004). It is found that it is almost impossible to operate a robotic task depending only on a single video feed that shows the cell from a single angle.

While investigating the learning process of robotic cells operation, it is found that the 3D view of the cell is a vital ingredient for the success of the learning process. A common way for teaching robotic operation is through 3D simulation programs, enabling the users to learn and experience with a variety of simulated work cells, some of which differ from the existing work cells in the lab. The next step was to find a way to combine the capabilities offered by the simulation program with basic teleoperation tools, creating not only a remote control interface but an actual remote learning mechanism.

Goldstain et al. (2007) present a schematic integration between Remote-Learning and teleoperation, significantly improving the learning process of the users and yielding an efficient use of equipment and resources. This integration enables the user to participate in a remote learning process on top of being able to operate a robotic system. The added value of such integration is in the ability to apply it to the available equipment and to different lab experiments (Michau et al., 2001; Ammari and Ben-Hadj, 2006).

The Compared Interfaces

Two different design platforms were used throughout the experiment: an INTERNET (Web-based) platform based on our Test Oriented Interface (TOI), as detailed later in the chapter, and a (wired) Robocell platform, which was operated either remotely or locally for comparison purposes. More specifically, four different interfaces were tested for the evaluation of teleoperation tasks:

- **INTERNET:** a Test Oriented Interface operated remotely (based on our methodology)
- **LRC:** a Robocell interface operated locally
- **VRC:** a Robocell interface operated remotely with Virtual Real-Time Presentation (VRTP)
- **RRC:** a Robocell interface operated remotely without VRTP

It is important to note that although the LRC is included within the compared interfaces, it is not a remote interface setting. This setting represents the everyday hands-on execution of robot operation locally in the laboratory and is used here as a benchmark to compare to the other settings, enabling evaluation of their proximity to hands-on operation performance.

The compared interfaces are described in the next sections.

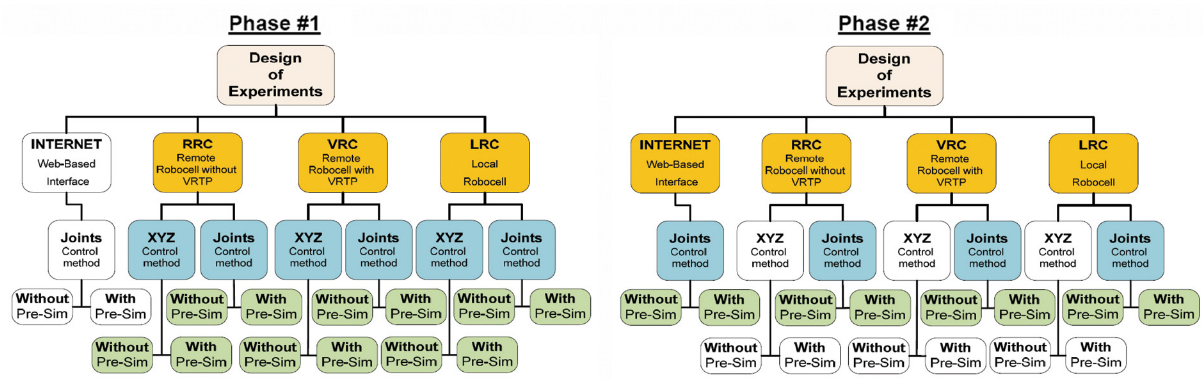
The Design of Experiments

As not all of the factors were technically able to be integrated into all the emulated interfaces, a partial factorial design was generated to include the available combinations of factors that could be tested. The partial design is motivated by practical and methodological considerations, as each interface was introduced with slight variations of the components for practical reasons. To avoid partial designs, the evaluation tests were divided into two congruent phases, each phase was evaluated as a full factorial model on its own, and together they covered all the design variations that were technically available.

Figure 2 describes the experimental tree for the design phases. The transparent branches represent the excluded parts of the experiment in that phase (Phase 1 lacks the Axis-XYZ control parameter and, in Phase 2, we aim at maintaining a balanced hierarchical experiment that includes the internet interface).

The first evaluation phase examines (1) the effect of the control method on the execution and the learning process and (2) the contribution of the VRTP module (integrated in the VRC interface). The second evaluation phase examines (1) the difference between the different RoboCell interfaces (LRC, RRC and VRC) and the designed remote Internet interface and (2) the contribution of the preliminary simulation tool in all the remote interfaces (Internet, RRC, VRC).

Figure 2. The two evaluation phases of the experiment



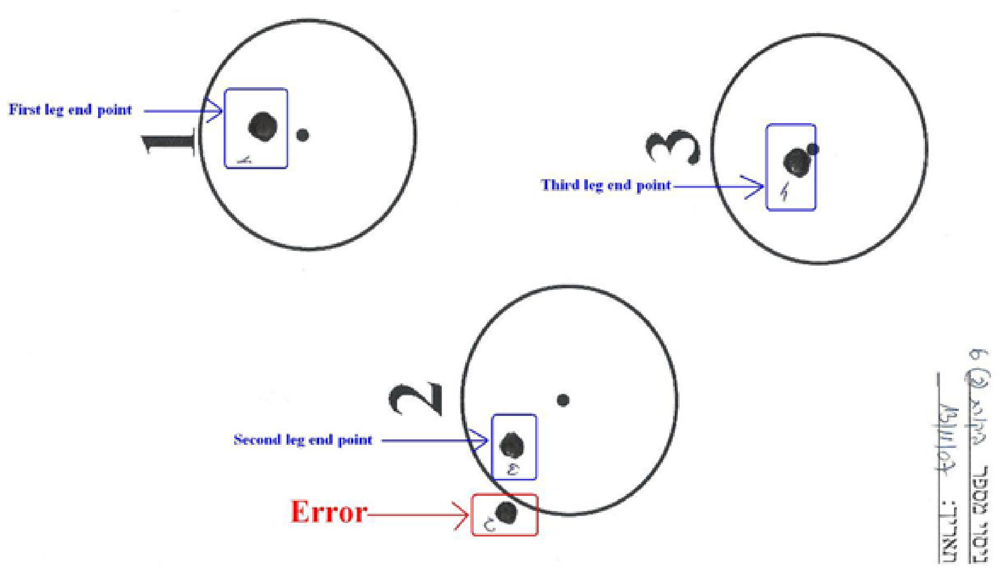
THE EXPERIMENT: INTERFACE COMPONENTS ANALYSIS

This section describes the experimental study. The goal of the experiment is to evaluate the effect of different interface components on the usability of the remote learning interface. This section presents the task and the performance measures and main design factors.

The Experimental Task

The designed task for the experiment was a simple "Pick-and-Place" task. It was adjusted to suit a remote operated system in the following way: the users were instructed to manipulate a robotic arm with a marker attached to its gripper in a way they will reach and mark a dot within three pre-placed circles. Figure 3 shows an actual result page of an experiment, describing the three pre-placed circles and the actual marked points.

Figure 3. An example of a completed worksheet



As the learning goal of the task is an efficient operation of the robot while using a remote interface, the users were instructed to try to perform the task in as an efficient way as possible, i.e. in the least number of movements possible, and with as few as possible markings outside the designated circles. The circles were placed in the exact same location in all different variations of the experiment. The starting point was defined as the "homing point" of the robotic arm.

Three performance measures were considered in the experiments, as discussed next.

Number of Movements Required to Complete a Leg

In the experiment, a user step (movement) was defined as a single press on one of the controller's buttons. In the examined designs, the movement was recorded as long as the button was pressed and was stopped when the button was released. Obviously, several movements were required to move the robot's arm from point to point. Every single movement was recorded on a data-recording sheet and was associated with one of the legs. A leg was defined as the period between reaching and marking each of the circles. Thus, the first leg includes all movements recorded from the homing point until marking the first circle; the second leg was defined between marking the first circle and the second circle, and so on.

Such steps assessed the overall performance of an operator and provided quantitative input regarding the improvement in performance during the whole task execution.

Number of Errors Recorded

Errors were defined as a marking made outside of the designated circles (for example, see the indicated error in Figure 3). Every mark on the worksheet was numbered and recorded, and the number of markings outside the circles was later analyzed. The number of errors provided information regarding the complexity of the task or the settings and helped to evaluate the performance of various interface designs.

Improvement Measure (Learning Curve)

An improvement measure was evaluated by using the number of movements measured in the legs. The learning rate (termed for simplicity reasons) was defined as follows:

$$LR = (N_{total} - N_{H_{to}1}) / N_{total}$$

where N_{total} is the total number of movements and $N_{H_{to}1}$ is the number of movements in the first leg (from the homing point to first circle). Dividing the number of movements required for legs two and three by the number of total required movements was used as an approximate criterion to assess the improvement and the learning pace during the execution. The lower this ratio was, the larger the part of the total movements was required by the first leg. In such a case, the subsequent two legs required significantly fewer movements, and therefore, one could conclude that the improvement from the first leg to the next ones was greater.

Evaluated Design Factors

Three components were defined as main design factors and evaluated through a set of experiments. The design factors were (i) use of a simulation tool prior to the task's execution; (ii) use of VRTP during the robotic task's execution; and (iii) the type of control method for the robotic arm. Each of these factors is presented next.

Preliminary Simulation

The used 3-D simulation tool is based on the Robocell© software cell-setup module. This tool allows the operator to experience and learn in advance (i.e., in an off-line mode) the robotic system and its environment. It provides the operator with a virtual working cell, similar to the one in the actual site, viewable from all directions and angles. It integrates the SCORBASE© robotic control software with interactive three-dimensional solid modeling simulation software in that it replicates the actual dimensions and functions of the real equipment, providing users with a fully simulated robotic learning environment and a graphic tracking view of the robot's operations.

This preliminary simulation tool was available in all the considered interfaces. It was expected to improve the execution of the task and to shorten the learning period required during the online operation of the robotic cell. For analysis purposes, the different interface settings were examined both with and without the use of preliminary simulation. To support a transfer of the practice from the simulation to the real operation, the preliminary simulation manual control modul was designed as similar as possible to the tele-robotic control interfaces.

Virtual Real-Time Presentation (VRTP)

In certain settings, the 3-D simulation tool can be operated not only in advance but also during the online execution of a task. When operated in an online mode, we refer to it as VRTP. The VRTP tool provides the operator with an extra view of the working area, including the possibility of changing the viewpoint in direction and orientation during the actual execution.

While this tool seems unnecessary when operating the robotic cell locally via eye contact, it may provide valuable information for a remote operator. The reason is that the remote operator receives 3D information on a 2D media channel, which may cause a misinterpretation or misperception of the environment. Unfortunately, due to its technical complexity, this tool is not integrated in most of the currently used remote learning interfaces. For this reason, it was not integrated into the TOI. Instead, for analyzing the contribution of this tool, we considered two different remote designs: one with and one without the VRTP tool, and called them the VRC and RRC interfaces, respectively.

Control Method of the Robotic Arm

Two common methods are available for controlling a robotic arm: the Axis-XYZ control method and the Joints control method.

The Axis-XYZ control method allows the operator to move the robotic arm along the axis of an imaginary Cartesian workspace. The linear movement is intuitive, but it requires greater computing resources as the robot's controller needs to calculate the exact direction and force for each joint motor and operate multiple motors simultaneously to achieve a linear movement. For these reasons, it was technically impossible to implement the required matrix into the TOI, and therefore, the Axis-XYZ control method was tested only in the LRC, RRC and VRC interfaces.

The Joints control method, although not as intuitive, is technically and mechanically much simpler. This control method is based on activating a single joint motor at a time, resulting in a nonlinear movement (in the case of a polaric joint). Consequently, it is more complicated for the inexperienced operator to control a robot using this method.

EXPERIMENTAL SETTINGS AND APPARATUS

Using the "web-based" design scheme, a dedicated website interface was designed to represent the Internet interface in the experiments. The website combined a remote controlling interface for the robot's arm along with a module for simulating and optimizing the operation, recording it, and downloading a pre-tested program to the robot's controller.

Physical Layout

A dedicated remote workstation was assembled to support the experiments. The workstation was equipped with two screens to standardize the visual feedback size and its position for all the experimented settings. The first screen was used for visual feedback only, displaying two live video feeds of the robotic cell: an overall view - an isometric view of the cell, and a zoom-in top-view of the work area. The second screen was used to support the actual teleoperation of the robotic cell. Both the Robocell software and an Internet browser were installed in the work station and alternately operated, depending on the experimental stage. Another workstation was used in the local site to host the servers of the website and the live video feeds.

Four interface versions were examined, dictating four slightly different settings for the local and the remote stations, as explained next.

Local Robocell (LRC) Interface

The local Robocell was used as a control group in the experiments. The LRC was executed right next to the robotic cell itself. In these experiments, the Robocell software was executed without adding any visual feedback. Since the experiments took place within eye-contact distance of the cell, it provided the operator with the opportunity to actually observe the robot and decide on the next required step.

Remote Robocell (RRC) Interface

The Remote Robocell workstation was equipped with a dedicated software package that was installed on the remote teleoperation computer and supported the preliminary simulation module. The teleoperation computer itself was wire-connected to the robot's controller via an amplified USB port. The remote workstation was placed in a remote room, preventing any eye contact between the operator and the robotic cell. These interface settings were designed for the examination of a modern teleoperation system that can support both the Axis-XYZ control and the joints control methods.

When using the RRC settings during the experiment, the operator had a presentation of the manual movement module of the Robocell software on one screen and two live video feeds of visual feedbacks on the other screen. In experiments where the preliminary simulation was used, the software was initially set to the off-line working mode. This means that the simulation module was presented to the users for an unlimited time, according to their choice, before turning into the online operation itself. Once changed to the actual online operation, the simulation module was shut down automatically.

VRTP-Enabled Remote Robocell (VRC) Interface

The VRTP-enabled remote Robocell mainly refers to the online option for simulation with the Robocell software.

The settings of both local and remote workstations were identical to those described for the RRC. The main difference between the two was the ability to keep the simulation module also running during the online operation stage. Such an option enabled the operator to have, on top of the visual feedback of the live feeds, a virtual visual feedback from the simulation module that was updated simultaneously with the movements of the robotic arm.

Potentially, such an option provided the users with both an advantage and a disadvantage with respect to the RRC. On one hand, they could change the orientation, angles, and zooming of the view in the VRTP module and gain better information than that obtained from the video feeds alone. However on the other hand, the virtual simulation could never be as accurate as the real video feed and could have resulted in operation errors, especially when the users reached the edge of the robot's working envelope.

INTERNET Interface

The Internet workstation contained, in addition to the standardized visual feedback mentioned above, a web browser. Setting only the internet browser was used to execute the teleoperation at the remote site.

The web browser was used to log into the proposed TOI and then to remotely operate the robotic arm through it. The visual-feed module in the TOI was not operated during the experiments to keep the same visual feedback for all interfaces, as required for analysis purposes. Instead the operator was provided with a separate visual feedback on the second screen, as happened in all the other remote settings. If a preliminary simulation had to be used in the experiment, the Robocell software was initially operated on the same computer, but only in an offline working mode, and the simulation module was presented to the operator for an unlimited time before turning into the online operation through the website. Once changing to the website TOI online operation, the Robocell software was shut down.

The Subjects Group

The subject group for the conducted experiments were senior (fourth year) students, at the CIM Lab in the IE Department at Tel-Aviv University. Each subject performed only a single experiment, randomly allocated to him. Each valid permutation of the design (14 in total) was performed by nine different subjects, and the results shown later are a mean over these repetitions. The subjects' age range lies from 20 to 30. The gender distribution was 53% males and 47% females. All subjects had a technical background resulting from their engineering education. Overall, 126 experiments were conducted throughout five semesters. The selection of this field of subject is obvious based on the results of this work, as engineering students are most likely the target end users of any system that might be developed.

EXPERIMENTAL RESULTS

Findings from Phase 1 of the Experiments

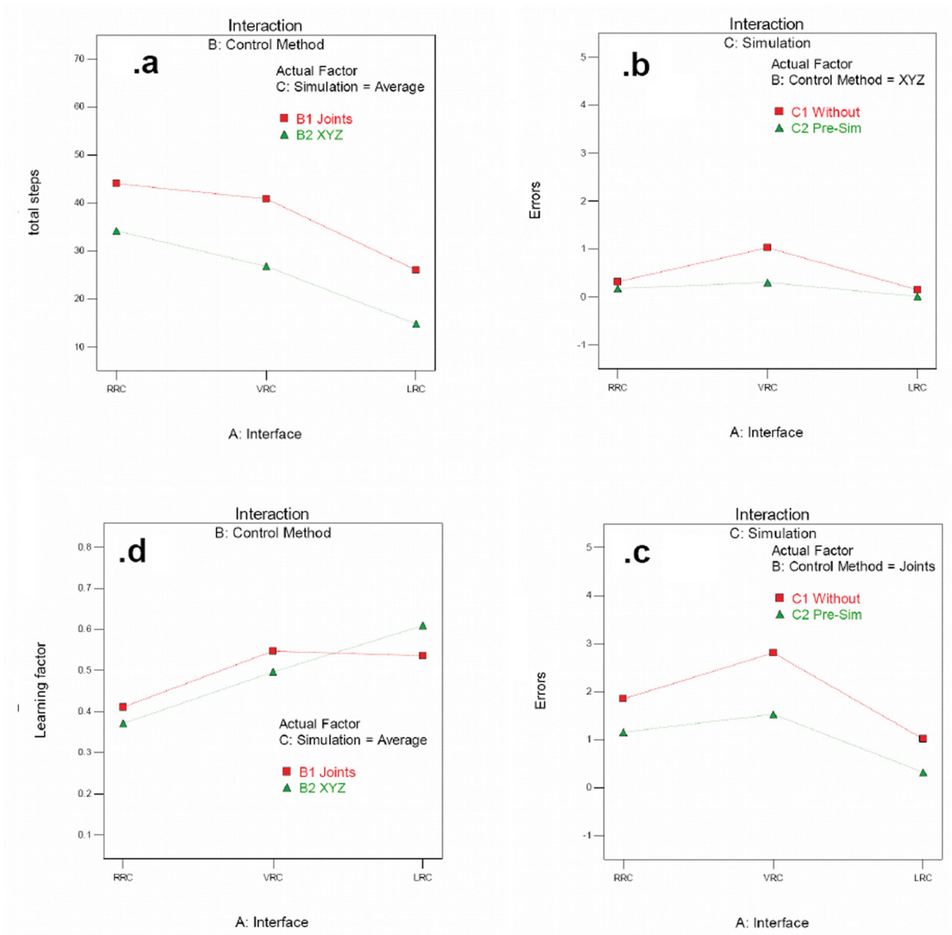
Phase 1 of the experiments focused on the evaluation of the three Robocell interfaces: the local Robocell interface (LRC), the remote Robocell interface that includes the VRTP (VRC), and the remote Robocell interface without the VRTP (RRC). As indicated earlier in the chapter, the evaluation focused on the XYZ vs. the Joints control method and on the contribution of the preliminary simulation tool prior to execution of the task.

The Total Number of Required Steps

The total number of steps required to complete the task was measured from the homing point (the location of the gripper after the calibration of the robot) until the completion of the entire task, i.e., after marking the third circle. This measure provided information and an intuitive understanding regarding the complexity level of the used interface.

Figure 4.a presents the average total number of steps required in each of the three Robocell-based interfaces, (starting from the RRC on the left hand side of the chart, continuing with the VRC in the middle and ending with the LRC on the right) and for each of the control methods. Note that as the interface approaches a realistic local setting, the number of steps decline. One can clearly see that the Joints control method results in a significantly

Figure 4. Selected graphs from phase 1 of the experiments



larger number of steps compared to the axis-XYZ control method. This trend is consistent through all the tested interfaces. This result is quite intuitive, as the Axis-XYZ control method is less complex and is more intuitive.

Examining the influence of the preliminary simulation tool on the total number of required steps, one can see that the preliminary simulation often leads to lower number of steps.

Number of Errors During the Tasks' Execution

The number of errors was measured by counting the number of marks made by the operator outside the designated circles in the working area. As mentioned for the previous factor, this factor gives some indication regarding the complexity of the task and points to the expertise improvement of the operator while completing the task. We expect the number of errors to be lower in designs that support better learning, as the user is expected to adapt faster to better control over the system and therefore to perform more accurately and with fewer errors.

The graphs in Figures 4.b and 4.c present again the average number of execution errors. This time, squares represent settings without a preliminary simulation and triangles represent settings with the use of a preliminary simulation.

Figures 4.b and 4.c explore the interaction between the control method and the preliminary simulation module when using the axis-XYZ or the Joints control, respectively. These show the effect of the control method used by the operator on the number of errors. As expected, the axis-XYZ control method reduces the number of errors with respect to the Joints control method and for all the examined interfaces in this phase. Surprisingly, the highest number of errors was obtained in the VRC interface rather than the RRC interface (the one without the VRTP). This result can be explained by the attractiveness of the VRTP feature, causing the users to rely on the virtual feedback even when it causes errors. The virtual feedback (VRTP), as informative as it is, is not as accurate as the online video feedback. When using the RRC interface without the virtual feedback, the operator had to wait for the video buffering delay to end, and therefore every movement took a longer time to finish. A possible explanation is that the operators gave a higher attention to each robotic move before actually executing it. We assume that the extra time and attention led to fewer errors.

Note that for the local-LRC interface, the advantage of the axis-XYZ with respect to Joints is smaller than the advantage in the remote interfaces. In fact, for the LRC interface, the control intervals overlap in contrast to the other interfaces. This fact emphasizes the importance of the control method when designing a remote interface.

While using the axis-XYZ control method (Figures 4.c), the average number of execution errors for both the local LRC interface and the remote RRC interface (without the VRTP) are found to be almost indistinguishable, regardless of the use or absence of the preliminary simulation module. In the VRC interface, on the other hand, a lower number of execution errors is obtained when implementing a preliminary simulation.

Note that for the Joints control method, the use of preliminary simulation results in a lower average number of errors in comparison to a situation without the preliminary simulation. This observation is consistent throughout all three different interfaces.

In both Figures 4.b and 4.c, one can see that the contribution of the preliminary simulation tool to decrease the number of execution errors is affected by the chosen control method. Preliminary simulation results in better execution (hence, better learning) when using an interface operated by the Joints control method.

The effect of the preliminary simulation on the number of errors while using the VRC interface seems to be almost indifferent to the control method in use. In this case, preliminary simulation results in a lower number of errors for both control methods, speculating that the VRTP module reduces the complexity gap between axis-XYZ and Joint control methods.

Learning and Improvement Measures

The learning and improvement rate, as defined earlier, is calculated by dividing the total number of steps required to perform the second and third legs, by the total number of steps required to complete the task. Such a measure is expected to be lower for a higher learning rate. A lower measure will indicate better learning, as it represents a significantly higher number of movements for the first leg and, therefore, a greater margin between the legs. This will indicate improved performance and an efficient learning process.

Figure 4.d presents the average learning function as calculated for the three interfaces that were considered. When using the local LRC interface, the Joints control method resulted in a better (lower) learning factor. This result is explained by the simplicity of the task when performed locally and with the simplest control method (axis-XYZ). A simple task leaves very little room for improvement, as its execution requires almost the minimal number of steps possible, those already from the first leg.

When using the remote interfaces (RRC and VRC), the axis-XYZ control method results in a slightly better improvement rate than the Joints control method. The best improvement and learning rate is obtained for the RRC interface, either with the axis-XYZ or with Joints control method. This result can also be explained by the complexity of the task to the remote users, leaving much room for learning and improvement.

Findings from Phase 2 of the Experiments

Phase 2 of the experiment focuses on the evaluation of the differences between the Web-based remote interface (INTERNET) and the three Robocell-based interfaces: the local settings (LRC), the remote settings with the VRTP module (VRC) and the remote settings without the VRTP (RRC). All four interface settings were operated with the Joints control method to evaluate the effect of the preliminary simulation tool prior to the online execution of the task. The results of Phase 2 are demonstrated in Figure 5.

The Total Average Number of Required Steps

Figure 5.a addresses the interaction between the interface type and the use of a preliminary simulation tool. When considering the three remote interfaces (the three most left interfaces in the abscissa), we see that the effect of the preliminary simulation to decrease the average total number of steps is most significant for the INTERNET interface. Such an effect barely exists in the VRC interface. Yet, the preliminary simulation also affects the total number of steps when using the local LRC interface.

Out of these three remote interfaces, the VRC interface is surprisingly indifferent to a preliminary simulation (as observed in Phase 1 of the experiment) and results in the same average number of steps both with and without using the preliminary simulation tool. Such a result may be explained by the fact that using the advanced VRTP tool during the task's execution makes the preliminary simulation, which is based on the same simulation tool, redundant regarding minimizing the number of steps. However, as explained in the next subsection, the preliminary simulation is useful regarding minimizing the number of errors.

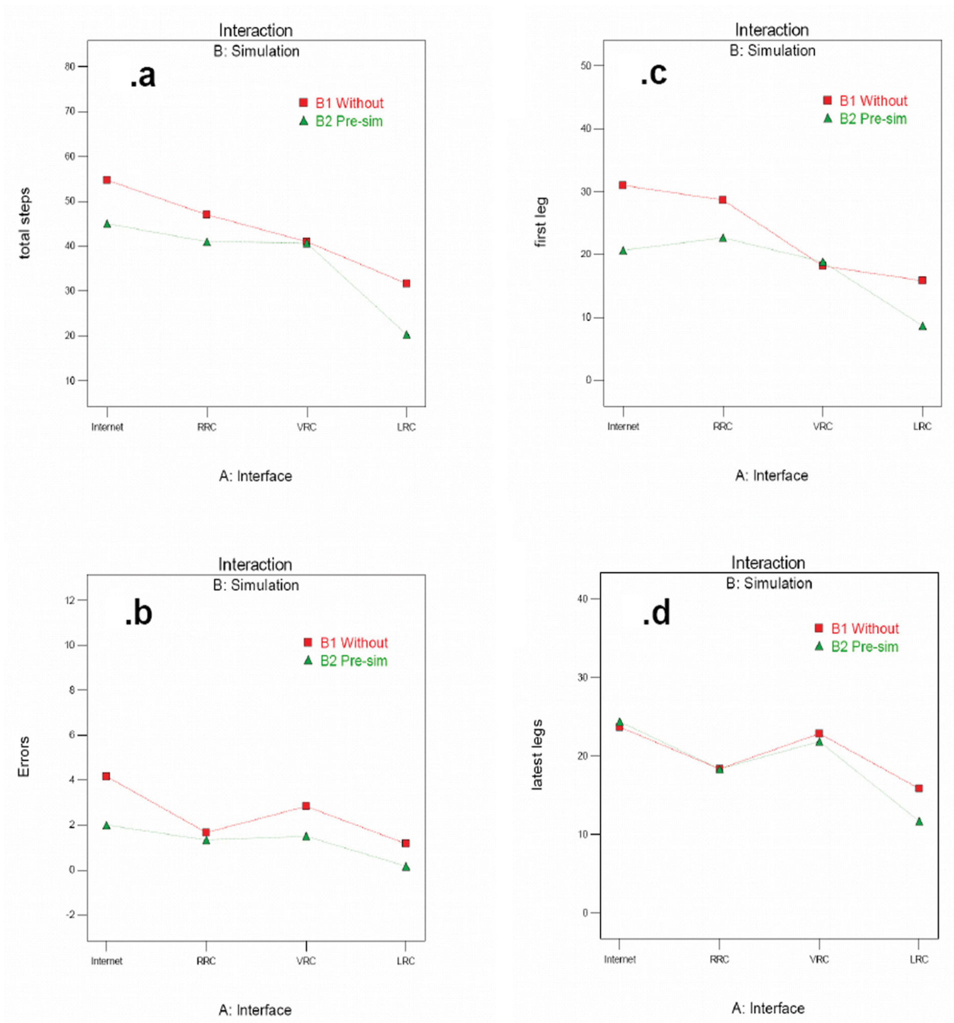
The Number of Occurred Errors

Figure 5.b presents the average number of execution errors for each of the four interfaces with or without the preliminary simulation tool. One can see that when using the preliminary simulation (marked by triangles), the difference between the three remote interfaces is negligible (less than 0.5 errors in average). Moreover, when the preliminary simulation is not used (marked by squares), the remote RRC interface lead to less errors than the other remote interfaces, obtaining an error rate, which is close to the one obtained by the local LRC test interface.

Learning and Improvement Measured

Next, the improvement in the learning rate of the users is presented. We show the learning rate for each of the legs as a function of the interface type and the use of a preliminary simulation. Figures 5.c and 5.d present the number of steps required for the first leg (marked as “H-to-1”) and for the later legs (marked as “1-to-3”), respectively. The results in each graph separately represent the use (by squares) or the lack of use (by triangles) of the preliminary simulation. As seen from the total number of required steps (Figure 5.a) and from the learning graphs (Figure 6), the use of a preliminary simulation results in fewer steps required to complete the first leg for all interfaces except for the VRC interface. Figure 5.d shows that the number of steps required to complete the later legs is indifferent to the use of a preliminary simulation, i.e., resulting in a roughly similar performance that is measured both with and without the use of the preliminary simulation tool prior to the execution.

Figure 5. Selected graphs from phase 2 of the experiments



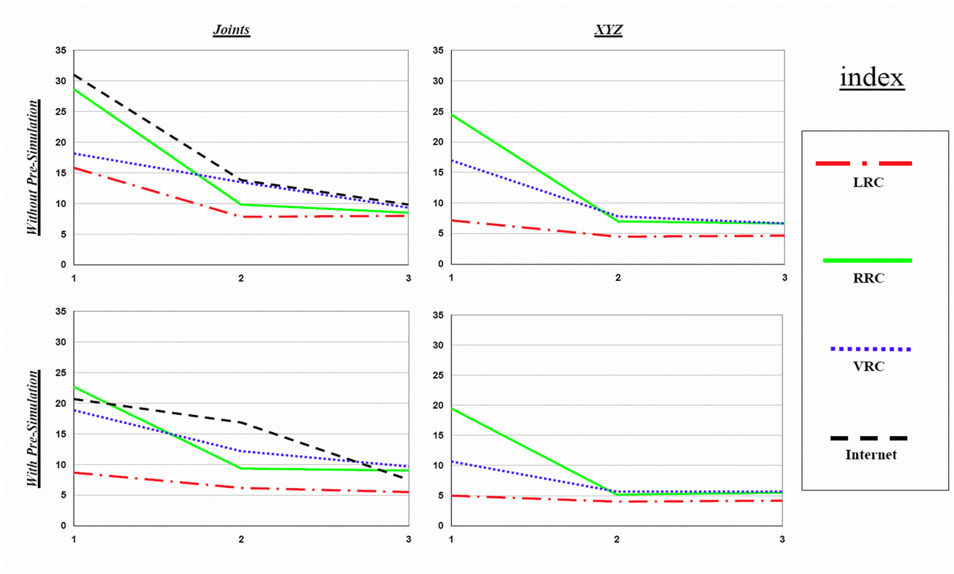
These results suggest that the ability to simulate movements in the cell before an actual execution provides the operator with an early stage of learning. This results in an improved performance at the beginning of the execution. The effect of such learning diminishes in later stages when the operator gains experience in remote operation. Another interesting result that is evident from these two graphs is that when a VRTP module is available, the contribution of the preliminary simulation is limited. Therefore, the use of VRTP can save time when the task is learned.

Learning Curves for Different Designs

The graphs shown in Figure 6 emulate the learning curves of the task execution for four combinations of the examined factors.

The horizontal axis represents the three legs of the experiments, starting from the homing point at the left hand side, through to the end of the execution on the right. The vertical axis shows the number of steps required for the specific leg of the execution. The right hand-side graphs show three different curves differentiated by the control

Figure 6. Learning curves for the experiments



method used and by the chosen interface (Internet interface excluded). The left hand-side graphs are showing four different curves differentiated by the use or absence of preliminary simulation and by the chosen interface (all four types of interfaces are included).

Comparison within Control-Method Factor

The two graphs on the right side of Figure 6 present the results of Phase 1 of the experiments. In each examined interface, both the Joint and the axis-XYZ control methods result in very similar slopes of the learning curve differing only in their heights. These differences result from the heights of start points, which are lower for the (more intuitive) axis-XYZ control method. The improvement throughout the steps seems to be almost unaffected by the actual control method in use. However, we see that while the end-points of all curves are bounded by a narrow margin between four to ten steps required to complete the last leg, the curves have greater margin at their starting points. This result leads to the conclusion that the actual improvement of execution achieved during the task is greater in the RRC interface and the lowest in the local LRC benchmark interface.

Analyzing the Preliminary-Simulation Factor

The two graphs on the left side of Figure 6 present the results of Phase 2 of the experiments. Unlike the right side graphs, the learning-curve slopes in each interface are different when considered with or without a preliminary simulation. For both the RRC and the local LRC interfaces, we observe steeper curves for experiments without preliminary simulation, indicating that the actual improvement of execution achieved during the task is greater when operating the system without a preliminary simulation. Nevertheless, we note that the curves of each interface end closely to each other, as the users reach the same average number of steps in the last leg of the execution, regardless of the use of a preliminary simulation. This observation leads to conclude that preliminary simulation supports the learning prior to the tasks' execution, thus it is recommended for use.

These conclusions are also supported by the learning curves of the Internet interface. In this more difficult working interface, the difference between the curves is noticeable, suggesting that when using a preliminary simulation, most of the learning occurred during the simulation part and leaving little room for improvement and providing a better start point to the operators at the online stage.

The curves for the VRC interface support our assumption that the VRTP tool and the preliminary simulation are superfluous to each other. This is supported by the almost identical curves (both in slope and in height), indicating indifference to the presence or absence of the preliminary simulation in the process.

Although the Internet interface is the most complex interface among the considered ones, when combined with a preliminary simulation tool, it provides almost as good results as the rest of the remote interfaces (in terms of the number of steps required). The same conclusion is drawn with respect to the number of errors. We believe that the ability to reconsider a movement once it has been chosen, yet before it is executed has a significant influence when explaining the success of this interface.

GUIDELINES AND FINAL DESIGN

When required to design a remote teleoperation interface, we need to choose the appropriate combination of components to meet our learning/teaching goals. If the goal is, as in our study, an accurate operation, then the suggested control method should be the axis-XYZ control, as it leads to a lower number of errors. However, since sometimes part of our robot operation teaching would benefit from teaching alternative control methods and as it seems that selecting either control method will not affect the achieved learning rate of the user, it is suggested to have both control methods available if and whenever it is possible.

A preliminary simulation module is highly recommended on the design of a remote telerobotic interface. The only module that was found to have the same impact as the preliminary simulation tool in terms of improving the operators learning and performance was the VRTP module. The VRTP module provides the user with the same learning qualities as the preliminary simulation tool, but this time during the actual online work. If the designed system has to service a large number of users by relying on short online time windows for each user, then a preliminary simulation is the most effective tool for learning. However, if one can provide each user with enough online access time to the robotic cell, then it is recommended to integrate a VRTP module into the interface.

A main feature that differs in the Internet interface from the other considered interfaces was the ability to reconsider a movement prior to its execution. We believe that this feature affected the higher learning rate found for this interface and recommend facilitating such mechanism into future designs of remote telerobotic interfaces.

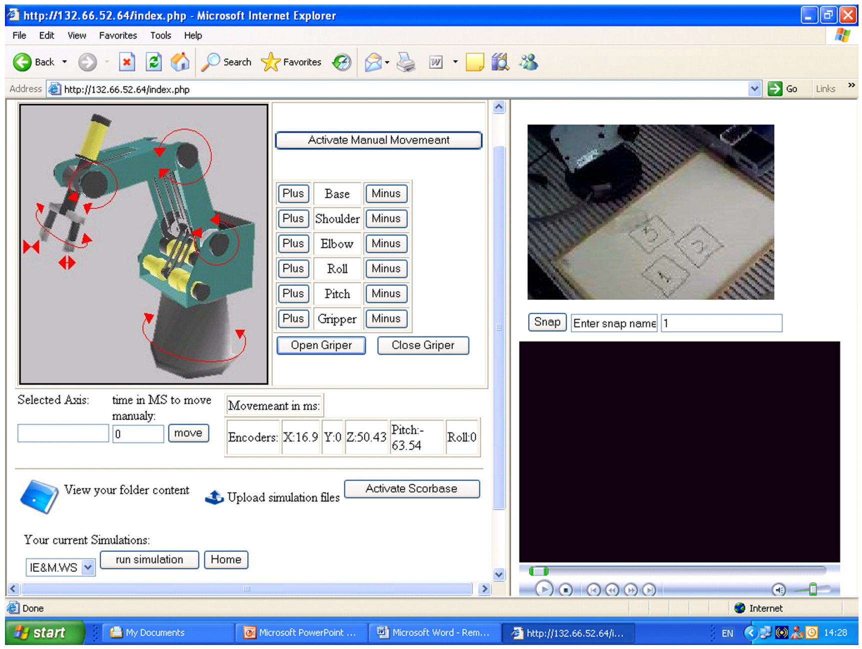
The design of the revised TOI designed based on the presented research is now presented.

Proposed Implementation

The web-based interface, shown in Figure 7, was programmed using html and PhP programming languages and was stored on an Apache server on the local computer (similar to Yang et al., 2004). The designed interface utilizes the Robocell simulation software as a platform.

Users were administrated with an SQL database, and the sessions were limited in time to avoid blockage of the system. The code for a single movement was based on measuring the time a specific push-button was pressed in the control module, identifying the joint and the direction represented by this button and then sending an operation command through the server and to the robot's controller. The response from the controller, composed of the new encoder values,

Figure 7. The web interface (© 2007, Taylor & Francis. Used with permission)



was presented on the website interface in response. A detailed description of the TOI modules is given next.

The Joint Control and Data Feedback Module

The Joint control and the data feedback module refer to the upper left hand-side of the interface, shown in Figure 7. This module controls directly each motor of the robot, enabling Joints control of the arm. Although the XYZ-axis control method was recommended as well for implementation, it does not appear in this design, as the implementation was too complex at the time of the actual design. A later work on the design should include the implementation of control method choice for the users.

There is a major difference between these two controllers and the Robocell interface controller described in previous sections. When using the Robocell controller, the action of clicking on a direction of movement results in an immediate response from the robotic arm, and the robot keeps moving in the desired direction until the push button is released. On the Internet interface presented here, pushing a controller button starts a timer (shown in Figure 7 under the Java application) running until the button is released. The operator can choose whether to send a movement signal to the robot for the selected axis and for that amount of time or to change the time/axis before sending the actual movement signal only after releasing the button. Such a feature potentially provides operators with greater control over their actions.

Once a movement was completed, a data feedback from the robot's encoders is presented to the operators (in the middle of Figure 7), enabling them to compare the actual positioning of the robotic arm to the one predicted by the virtual simulation.

The Visual Feedback Module

Two video feeds, as seen in the right hand side of Figure 7, are available for the users. These feeds provide the users with two different viewing angles of the workstation: an isometric overall view (the lower feed) and a zoomed top view (the upper feed). The top view also enables the users to snap a picture of the work area and to save it in their folder on the server. This option is introduced to support the maintenance of lab reports by future users of the system.

The Data Interaction Module

The bottom left side of Figure 7 shows the data interaction module serving three purposes: 1) uploading files to the server, 2) viewing the personal folder of the logged user, and 3) running simulation files stored in this folder. This module is not used during this research and was extensively described in Goldstain et al. (2007).

FURTHER RESEARCH

This chapter addresses the learning aspects related to the usage of different interface settings for remote control and manipulation of a robotic cell. Various interface components are evaluated and recommendations are provided.

Further research in this field can address the effect of integrating a VRTP tool and an axis-XYZ control method into an Internet interface. Results drawn from this research suggest that such integration can yield the best remote learning performance. Other research could focus on visual aspects associated with remote telerobotic learning, examining positioning and orientation of cameras and their effect on the user's comprehension of the three-dimensional work area as well as on their learning performance. In relation to teaching laboratories, useful work can be done for designing remote-compatible tasks for learning robotics, as not all available routines for teaching robotics are applicable for remote learning without the presence of an instructor on the site.

REFERENCES

Adams, J. A. (2002). Critical considerations for human–robot interface development. *In Proc. Human–Robot Interaction*, 2002 AAAI Fall Symp. Menlo Park, CA, 2002, (pp. 1-8).

Aktan, B., Bohus, C. A., Crowl, L. A., & Shor, M. H. (1996). Distance learning applied to control engineering laboratories. *IEEE Transactions on Education*, *39*(3), 320–326. doi:10.1109/13.538754

Ammari, A. C., & Ben Hadj Slama, J. (2006). The development of a remote laboratory for Internet-based engineering education. *Journal of Asynchronous Learning Networks*, *10*(4), 12.

Belousov, I., Chellali, R., & Clapworthy, J. (2001). virtual reality tools for Intemet robotics. *In Proc. of the ZOO1 IEEE Internationnl Confererence on Robotics & Automation*, Seoul, Korea

Belousov, I., Tan, J., & Clapworthy, G. (1999). Teleoperation and Java3D visualization of a robot manipulator over the World Wide Web. *In Proc. Information Visuualisation* IV'99, July 1999 (pp. 543-548).

Bochiccio, M. A., & Longo, A. (2009). Hands-on remote labs: Collaborative Web laboratories as a case study for IT engineering classes. *IEEE Transactions on Learning Technologies, 2*(4).

Bukchin, J., Luquer, R., & Stubs, A. (2002). Learning in tele-operations. *IIE Transactions*, *34*, 245–252. doi:10.1080/07408170208928866

Calkin, D. W., Parkin, R. M., Šafaric, R., & Czarnecki, C. A. (1998). Visualization, simulation and control of a robotic system using Internet technology. In *Proc. 5th IEEE Int. Advanced Motion Control Workshop*, Coimbra, Portugal, 1998, (pp. 339-404).

Candelas, F. A., Puente, S. T., Torres, F., Segarra, V., & Navarrete, J. (2005). Flexible system for simulating and tele-operating robots through the Internet. *Journal of Robotic Systems*, *22*(3), 157–166. doi:10.1002/rob.20056

Chen, S. H., Chen, R., Ramakrishnan, V., Hu, S. Y., Zhuang, Y., Ko, C. C., & Chen, B. M. (1999). Development of remote laboratory experimentation through Internet. In *Proc. of the 1999 Hong Kong Symposium on Robotics and Control* vol. 2, (pp. 756-760).

Chiculita, C., & Frangu, L. (2002). A Web based remote control laboratory. In *Proc. Of 6th Multiconference on Systemic, Cybernetics and informatics, Orlando, Fl.*, 2002.

Cooper, M., Donnelly, A., & Ferreira, J. (2002). Remote controlled experiments for teaching over the Internet: A comparison of approaches developed in the PEARL project. *19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 119-128).

Doulgeri, Z., & Matiakis, T. (2006). A web telerobotic system to teach industrial robot path planning and control. *IEEE Transactions on Education*, *49*(2). doi:10.1109/TE.2006.873975

Eliav, A., Lavie, T., Parmet, Y., Stern, H., Waches, J., & Edan, Y. (2005). *Kiss human-robot interfaces*. 18th International Conference on Production Research (ICPR).

Enrique Sucar, L., Noguez, J., & Huesca, G. (2005). *Project oriented learning for basic robotics using virtual laboratories and intelligent tutors.* 35th ASEE/IEEE Frontiers in Education Conference, Indianapolis, IN, October 2005.

Fernandes, A. S. C., & Martins, M. J. (2001). Self-training through the Internet. *European Journal of Engineering Education*, *26*(2), 169–177. doi:10.1080/03043790110034429

Goldstain, O., Ben-Gal, I., & Bukchin, Y. (2007). Remote learning for the manipulation and control of robotic cells. *European Journal of Engineering Education*, *32*(4), 481–494. doi:10.1080/03043790701337213

Goldstain, O., Ben-Gal, I., & Bukchin, Y. (2010). *Evaluation of tele-robotic interface components for teaching robot operation.* Unpublished manuscript.

Gravier, C., Fayolle, J., Bayard, B., Ates, M., & Lardon, J. (2008). State of the art about remote laboratories paradigms Foundations of ongoing mutations. [iJOE]. *Int. J. Online Eng.*, *4*(1), 19–25.

Henry, J., & Schaedel, H. M. (2005). International co-operation in control engineering education using online experiments. *European Journal of Engineering Education*, *30*(2), 265–274. doi:10.1080/03043790410001065678

Hu, H., Yu, L. P., Tsui, W., & Zhou, Q. (2001). Internet-based robotic systems for teleoperation. *Assembly Automation*, *21*(2), 143–151. doi:10.1108/01445150110388513

Khan, W. A., Al-Doussari, S. M. R., & Al-Kahtani, A. H. M. (2002). Establishment of engineering laboratories for undergraduate and postgraduate studies. *European Journal of Engineering Education*, *27*(4), 425–435. doi:10.1080/03043790210166684

Kofman, J., Xianghai, W., Luu, T., & Verma, S. (2005). Teleoperation of a robot manipulator using a vision-based human-robot interface. *IEEE Transactions on Industrial Electronics*, *52*, 1206. doi:10.1109/TIE.2005.855696

Marin, R., Sanz, P. J., Nebot, P., & Wirz, R. (2005). A multimodal interface to control a robot arm via the Web: A case study on remote programming. *IEEE Transactions on Industrial Electronics*, *52*, 1506. doi:10.1109/TIE.2005.858733

Michau, F., Gentil, S., & Barrault, M. (2001). Expected benefits of web-based learning for engineering education: examples in control engineering. *European Journal of Engineering Education*, *26*(2), 151–168. doi:10.1080/03043790110034410

NASA. (n.d.). *Space telerobotics program*. Retrieved from http://rainer.oact.hq.nasa.gov/telerobotics_page/telerobotics.shtml

Puente, S. T., Torres, F., Ortiz, F. G., & Candelas, F. A. (2000). Remote robot execution through WWW simulation. In *Proc. 15th International Conference on Pattern Recognition*, ICPR 2000.

Saliah, H., Saad, M., Villardier, L., Assogba, B., Kedowide, C., & Wong, T. (2000) *Resource management strategies for remote virtual laboratory experimentation*. 2000 Frontier in Education Conference, Kansas City.

Siegwart, R., & Saucy, P. (1999). Interacting mobile robots on the web. In *IEEE International Conference on Robotics and Automation* (ICRA), 1999.

Tel-Aviv University. (n.d.). *CIM-Lab remote learning for robotic cells website*. Retrieved from http://www.eng.tau.ac.il/remote

Tzafestas, C. S., Palaiologou, N., & Alifragis, M. (2006). Virtual and remote robotic laboratory: Comparative experimental evaluation. *IEEE Transactions on Education*, *49*(3). doi:10.1109/TE.2006.879255

Wang, M., & Liu, J. N. K. (2004). A novel teleoperation paradigm for human-robot interaction. *IEEE International Conference on Robotics, Automation and Mechatronics*, (pp. 13–18).

Yang, X., Chen, Q., Petriu, D. C., & Petriu, E. M. (2004). Internet-based teleoperation of a robot manipulator for education. In *Proc. of the 3rd IEEE International Workshop on Haptic, Audio- and Visual Environments and Their Applications*, 2004, (pp. 7-11).

KEY TERMS AND DEFINITIONS

Home-Based: a design scheme for remote labs based on installing simulation and programming tools on the user's computer, sending batches of pre-planed code to be implemented on laboratory equipment.

INTERNET: A Test Oriented Interface operated remotely (based on our methodology)

Lab-Based: a design scheme for remote labs based on enabling direct remote access to laboratory equipment and computers.

LRC: A Robocell interface operated locally.

RoboCell: Robotic control software providing a user-friendly tool for robot programming and operation.

RRC: A Robocell interface operated remotely, without VRTP.

Scorbase: The language used by RoboCell's interface to program a robotic cell.

VRC: A Robocell interface operated remotely, with VRTP.

Web-Based: an integrated design scheme for remote labs based on a designated server running a pre-programmed website, detaching the user from the physical lab computer, while providing full control over laboratory equipment.

This work was previously published in Internet Accessible Remote Laboratories: Scalable E-Learning Tools for Engineering and Science Disciplines, edited by Abul K.M. Azad, Michael E. Auer and V. Judson Harward, pp. 163-184, copyright 2012 by Engineering Science Reference (an imprint of IGI Global).

Chapter 64
Cell Loading and Family Scheduling for Jobs with Individual Due Dates

Gürsel A. Süer
Ohio University, USA

Emre M. Mese
D.E. Foxx & Associates, Inc., USA

ABSTRACT

In this chapter, cell loading and family scheduling in a cellular manufacturing environment is studied. What separates this study from others is the presence of individual due dates for every job in a family. The performance measure is to minimize the number of tardy jobs. Family splitting among cells is allowed but job splitting is not. Even though family splitting increases number of setups, it increases the possibility of meeting individual job due dates. Two methods are employed in order to solve this problem, namely Mathematical Modeling and Genetic Algorithms. The results showed that Genetic Algorithm found the optimal solution for all problems tested. Furthermore, GA is efficient compared to the Mathematical Modeling especially for larger problems in terms of execution times. The results of experimentation showed that family splitting was observed in all multi-cell solutions, and therefore, it can be concluded that family splitting is a good strategy.

INTRODUCTION

Cell Loading is a decision making activity for planning the production in a Cellular Manufacturing System (CMS) including more than one manufacturing cell. The products are assigned to the manufacturing cells where they can be processed. This assignment is done based on the demand, processing times and due dates of the products and the production capacity and capability of the manufacturing cells (Süer, Saiz, Dagli & Gonzalez, 1995 and Süer, Saiz, & Gonzalez, 1999). Family Sequencing is a task of deciding the order by which product families are processed in a particular cell as determined by the Cell Loading process. In this chapter, a product

DOI: 10.4018/978-1-4666-1945-6.ch064

family can be split and they can be sequenced in the same cell or different cells. Obviously, each time a new family starts in a cell, a new setup is required. Finally, Family Scheduling consists of determining start times and completion times of the product families and the individual products based on the family sequence established. Typically in a complex cellular system, we need to address Cell Loading, Family Sequencing and Family Scheduling tasks all in a satisfactory manner to obtain the desired results in terms of the selected performance measure.

In this study, we are considering minimizing the number of tardy job as the performance measure. Even though the problem has been observed in a shoe manufacturing company, it is applicable to many cellular systems. The products are grouped into families based on their processing similarity. On the other hand, products in a family might have different due dates. The overall objective of this chapter is to solve cell loading and product sequencing problem in such a multi-cell environment. To accomplish this, we propose two different approaches to tackle this complex problem namely, mathematical modeling and genetic algorithms. An experiment is carried out using both approaches and later the results are compared and a sensitivity analysis is also performed with respect to due dates and setup times.

BACKGROUND

Group Technology (GT) is a general philosophy where similar things are grouped together and handled all together. GT is established upon a common principle that most of the problems can be grouped based on their similarities and then a single solution can be found to the entire group of problems to save time and effort. This general concept has been also applied to the manufacturing world. This approach increases productivity by reducing work-in-progress inventory and improves delivery performance by reducing leadtimes, thus helping manufacturing companies to be more competitive. Thus, a Cellular Manufacturing System can be specified as an application of GT to the manufacturing system design (Askin & Standridge, 1993). Cellular Manufacturing System aims to obtain the flexibility to produce a high or moderate variety of low or moderate demand products with high productivity. CMS is a type of manufacturing system that consists of manufacturing cell(s) with dissimilar machines needed to produce part family/families. Generally, the products grouped together form a product family. The benefits of CMS are lower setup, smaller lot sizes, lower work-in-process inventory and less space, material handling, and shorter throughput time, simpler work flow (Suresh & Kay, 1998).

In this chapter, the performance measure used is minimizing the number of tardy products (n_T). If a product is completed after its due date, then it is considered as tardy product. If product is completed before its due date, then the tardiness for this product will be zero (early or on-time product). Therefore, tardiness for a product takes a value of zero or positive, $T_i = \max\{0, c_i - d_i\}$; where T_i is the tardiness for product (i), c_i is the completion time of product (i), and d_i is the due date for product (i). The number of tardy jobs is computed as $n_T = \sum_i^n g(T_i)$ where g(x) = 1 if x > 0, and zero otherwise.

The problem has been observed in a shoe manufacturing company where twelve product families have been already defined. There are multiple cells and the most critical component of each cell is the rotary injection molding machine. Even though Rotary Molding Machine is a single machine, scheduling shoes on that machine resembles to a parallel machine scheduling problem as it can hold multiple pairs of molds/shoes at any time. The Rotary Molding Machine is defined in detail in Section 3.

Several researchers focused on cell loading problem [Süer, Saiz, Dagli & Gonzalez, (1995) and Süer, Saiz, & Gonzalez (1999a)] developed simple cell loading rules to minimize number of tardy jobs. Süer, Vazquez, & Cortes (2005) developed a hybrid approach of Genetic Algorithms and local optimizer to minimize nT in a multi-cell environment. Süer, Arikan, & Babayigit (2008) and (2009) focused on cell loading subject to manpower restrictions and developed fuzzy math models to minimize nT and total manpower levels. A few works have been also reported where both cell loading and product sequencing tasks are carried out. Süer & Dagli (2005) and Süer, Cosner & Patten (2009) discussed models to minimize makespan, machine requirements and manpower transfers. Yarimoglu (2009) developed math model and genetic algorithms to minimize manpower shortages in cells with synchronized material flow. However, these work ignored setup times between products and families.

Some other researchers focused on group scheduling problem with a single machine or a single manufacturing cell. Nakamura, Yoshida & Hitomi (1978) focused on minimizing total tardiness and considered sequence-independent group setup. Hitomi & Ham (1978) also considered sequence-independent setup times for a single machine. Ham, Hitomi, Nakamura & Yoshida (1979) developed a branch-and-bound algorithm for the optimal group and job sequence to minimize total flow time with the minimum number of tardy jobs. Pan and Wu (1998) considered a single machine scheduling problem to minimize mean flow time of all jobs subject to due date satisfaction. They categorized the jobs into groups without family splitting. Gupta and Chantaravarapan (2008) studied the single machine scheduling problem to minimize total tardiness considering group technology. Individual due dates and independent family setup times have been used in their problem with no family splitting.

This paragraph summarizes the work done in the past which focused on scheduling jobs on the Rotary Injection Molding Machines. Süer. Santos, & Vazquez (1999b) have developed a three-phase Heuristic Procedure to minimize makespan in the Rotary Molding Machine scheduling problem. Subramanian (2004) has attempted to solve this problem as a part of the cell loading and scheduling process. The objective was to minimize makespan and unlimited availability of the molds was assumed. Later, Urs (2005) introduced limited mold availability into the problem for the same objective. The most recent research was done by Süer, Subramanian, and Huang (2009) includes some heuristic procedures and mathematical models for cell loading and scheduling problem.

The most important feature of the scheduling problem studied in this chapter is the presence of individual due dates for every job even in the same family (no common due dates), and family splitting is allowed to minimize the number of tardy jobs. To the best knowledge of the authors, this real problem observed in a cellular environment has not been addressed in the literature before. As a result, we decided to tackle this complex problem here and propose multiple solution approaches to deal with it.

THE PROBLEM STUDIED

This section discusses the problem studied in detail.

Family Splitting vs. Setup Times

Typically, in cellular manufacturing similar products are grouped together and processed together as a family to reduce the number of setups and thus total setup time. However, literature does not address the possibility of having different due dates for the jobs in the same family. Even though it is important to reduce the setup times, it is also important to meet the customer due dates. There is a natural conflict between meeting due dates of jobs versus reducing total setup times between families. This point can be illustrated

Figure 1. Family splitting not allowed vs. allowed

	Setup Fam1	Job1 Fam1	Job2 Fam1	Setup Fam2	Job3 Fam2	Job4 Fam2
ci	2	5	10	12	16	22
di		6	19		12	20
Ti		0	0		4	2

a) Family splitting not allowed

	Setup Fam1	Job1 Fam1	Setup Fam2	Job3 Fam 2	Setup Fam1	Job2 Fam1	Setup Fam2	Job4 Fam2
ci	2	5	7	11	13	18	20	26
di		6		12		19		22
Ti		0		0		0		4

b) Family splitting allowed

in Figure 1 where there are two families and two jobs in each family.

If we do not split (preempt) the families, we get two jobs in the second family. On the other hand, if family splitting is allowed then number of tardy jobs is reduced to one even if the number of setups increases from two to four. We can observe that when all of the jobs for a family are scheduled all together, setup times are reduced. However, this may also force several other jobs in the following families to be delayed and increase the possibility of becoming tardy. On the other hand, when a family is split several times, the number of setups increases thus reducing the productive time and hence may adversely affect the number of tardy products. This study attempts to find a balance between family splitting and meeting due date such that the total number of tardy products is minimized. As mentioned before, among the published papers in the literature, there is no work reported about Cell Loading and Family Scheduling subject to Individual Due Dates with group splitting allowed considering more than one manufacturing cell.

Case Study

This section describes the problem in depth. The following subsections explain the important features of the problem.

Products

This problem was observed in a shoe manufacturing plant. Products have five attributes; Gender, Size, Sole Type, Color, and Material. For shoes manufactured for men (Male (M)), there are 18 different sizes, 2 sole types (Full Shot (FS), Mid Sole (MS)), 4 colors (Black (B), Dark Green (G), Honey (H), Nicotine (N)), and 3 materials (Polyurethane (PU), Polyvinyl chloride (PVC), Thermo Plastic Rubber (TPR)). For shoes manufactured for women (Female (F)), there are 13 different sizes and the remaining attributes (Sole Types, Colors, and Materials) are similar to those of males. Besides these product types, there are also different upper designs that will be referred as models from now on. Each model will have its own identification designation (Model ID).

Cells/Minicells

There are six manufacturing cells in the plant and they are independent from each other (machine sharing, and thus inter-cell transfers are not allowed). In the plant, every manufacturing cell includes Lasting Minicell, Rotary Molding Machine Minicell (RMMM), and Finishing/Packing Minicell as shown in Figure 2. Lasting Minicells prepare the shoes for injection molding process. Rotary Molding Machine Minicells inject the materials into the molds. Finishing/Packing Minicells remove extra materials from the injected shoes, finish the shoes, and also pack the shoes.

Rotary Molding Machine

This study focuses only on scheduling Rotary Molding Machine Minicells (the bottleneck of the manufacturing cell). The Rotary Molding Machine has a capacity of six pair molds as shown in Figure 3. In Figure 2, P1 is the injection station where the material is injected inside the mold. P2, P3, P4, and P5 are the cooling off stations, so that worker can handle the shoes. P6 is the loading and unloading station that the worker removes the pair injected and cooled off, and then loads the new pair that will be injected in the injection station. The Rotary Molding Machine is rotated anti-clockwise, so it is rotated exactly one position at the end of every cycle time.

Figure 2. Manufacturing cells in the shoe manufacturing plant

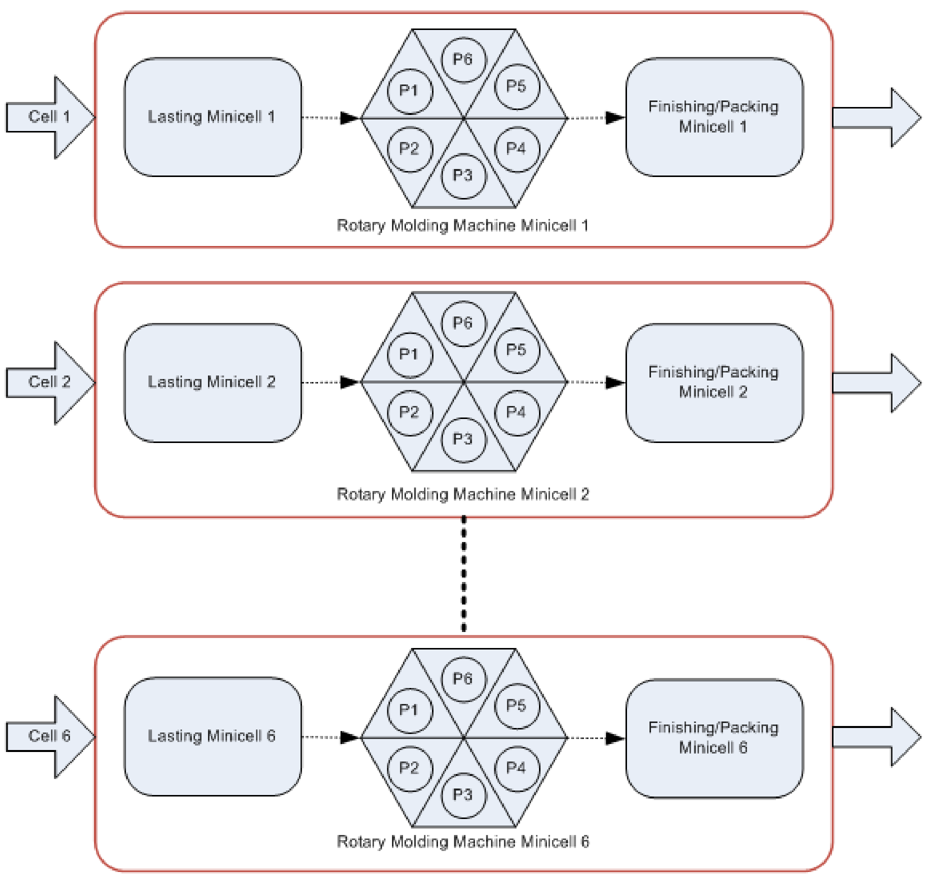
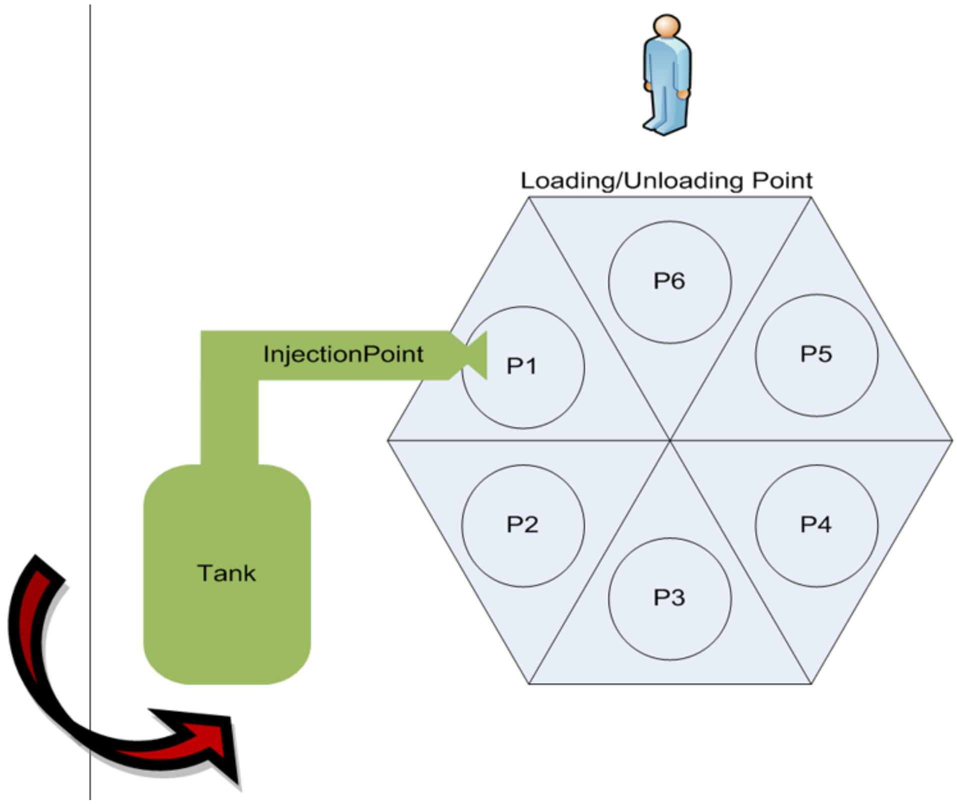


Figure 3. The rotary molding machine minicell



Injection time is defined as the time required for injecting the material inside the mold. The injection time is affected by the size of the shoe, i.e. larger shoe sizes need longer injection times, because the material injected by the Rotary Molding Machine per minute is constant. Because of the schedule of products in the specific cell, different sizes can be run in the Rotary Molding Machine at the same time. When this happens, the cycle time is set to the injection time of the biggest size (maximum injection time).

Product Families

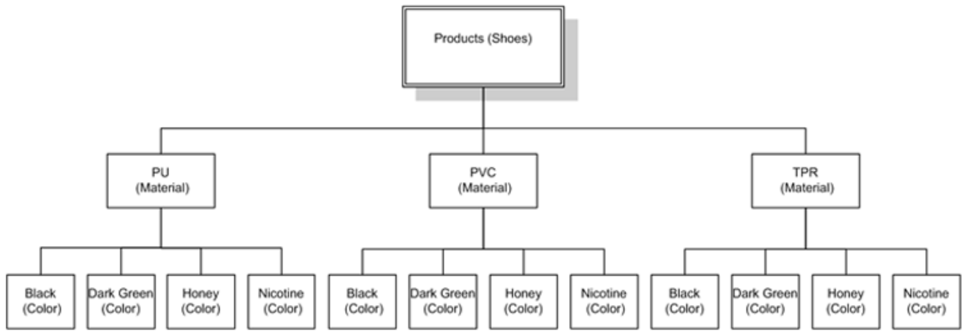
A representation code is formed as "MC" to form and identify the product families. In the MC code form: M denotes the Material (PU: U, PVC: P, TPR: T), and C denotes the Color (Black: B, Dark Green: G, Honey: H, Nicotine: N). There are 12 product families (= 4 colors * 3 material types) as shown in Figure 4.

In this study, all of the sizes of a specific order (with the same Model ID, Gender, Sole Type, Material, Color, and Due Date) is called a job. Different sizes of a job can have different demand. All of the sizes included in a job are assumed to have the same due date. The reason for this is that the entire job will have to be shipped to the customer all together.

Example: Family Formation

An example of customer orders that consists of 32 jobs is presented in Table 1. From the customer orders in Table 1, the families can be obtained as shown in Table 2.

Figure 4. Structure of families



Example: Possible Cases

This study focuses on assigning products to Rotary Molding Machine Minicells considering their families, i.e. cell loading. Family splitting among Minicells is allowed but job splitting is not. As an illustration of this, examples of the possible minicell loading cases are shown in Figure 4. The processing times of jobs in families F1, F5, and F12 are given in Table 3. While loading minicells, the families may not be divided as shown in Figure 5a, or families may be divided in the same minicell (like preemption) as shown in Figure 5b, or a family may be assigned to multiple minicells as shown in Figure 5c.

Other Issues

The molds used in the Rotary Molding Machine for injection molding vary by size, gender, and sole type. It is assumed that there is not any restriction on the availability of molds. Therefore, the same size pairs of a job can be run on all locations of the Rotary Molding Machine simultaneously. In this study, setup times between jobs in the same family are assumed negligible. However, setup times (for material or color or both changes) between families is assumed to take 20 minutes. The jobs can be back-scheduled in the Lasting Minicells and forward-scheduled in the Finishing/Packing Minicells based on the schedule of Rotary Molding Machine Minicells. However, scheduling in Lasting Minicells and Finishing/Packing Minicells are not within the scope of this work.

PROPOSED SOLUTION TECHNIQUES

The Cell Loading and Family Scheduling problems introduced in this chapter involve constraints on the number of product families, individual due dates, machine capacity, and sequence-independent setup times. This version of the problem is more difficult than the Classical Cell Loading and Group Scheduling problem. Both mathematical model and genetic algorithms approaches are proposed. The Mathematical Model guarantees the optimal solution, but it takes too much time to find the optimal solution. The Genetic Algorithm is a much faster procedure, but it cannot guarantee the optimal solution. The performance of these two procedures is compared with respect to execution time and the frequency of the optimal solutions. Genetic Algorithms Software Application (GASA) has been coded by using C# object-oriented programming language. The Mathematical Model is solved by using ILOG OPL 5.5. The methodology introduced in chapter is not restricted to the case study discussed here and can directly be used in similar cellular environments.

Table 1. An example of customer orders

Job No.	Model ID	Gender	Sole Type	Material	Color	Size	Code	Total Demand	Due Date
1	K	F	FS	TPR	Black	5, 6, 7, …, 12	TB	269	10
2	U	M	FS	PVC	Dark Green	5, 6, 7, …, 15	PG	688	12
3	C	M	FS	TPR	Black	5, 6, 7, …, 15	TB	1045	22
4	C	M	FS	TPR	Black	5, 6, 7, …, 15	TB	208	11
5	L	F	MS	PU	Black	5, 6, 7, …, 12	UB	881	20
6	T	M	MS	PU	Black	5, 6, 7, …, 15	UB	831	17
7	T	M	MS	PU	Black	5, 6, 7, …, 15	UB	277	13
8	O	F	FS	PVC	Dark Green	5, 6, 7, …, 12	PG	250	15
9	E	M	FS	PVC	Dark Green	5, 6, 7, …, 15	PG	636	11
10	W	M	FS	PVC	Black	5, 6, 7, …, 15	PB	384	14
11	W	M	FS	PVC	Black	5, 6, 7, …, 15	PB	329	16
12	F	F	MS	TPR	Dark Green	5, 6, 7, …, 12	TG	440	17
13	F	F	MS	TPR	Dark Green	5, 6, 7, …, 12	TG	321	11
14	N	F	MS	PU	Black	5, 6, 7, …, 12	UB	355	14
15	N	F	MS	PU	Black	5, 6, 7, …, 12	UB	255	10
16	X	M	MS	PVC	Black	5, 6, 7, …, 15	PB	788	20
17	E	F	FS	PVC	Nicotine	5, 6, 7, …, 12	PN	574	16
18	E	F	FS	PVC	Nicotine	5, 6, 7, …, 12	PN	245	12
19	Y	F	FS	PVC	Honey	5, 6, 7, …, 12	PH	456	14
20	G	M	FS	PVC	Honey	5, 6, 7, …, 15	PH	345	13
21	G	M	FS	PVC	Honey	5, 6, 7, …, 15	PH	657	16
22	O	M	FS	TPR	Honey	5, 6, 7, …, 15	TH	234	11
23	M	M	FS	PU	Nicotine	5, 6, 7, …, 15	UN	621	16
24	W	M	FS	TPR	Nicotine	5, 6, 7, …, 15	TN	206	12
25	P	F	MS	TPR	Nicotine	5, 6, 7, …, 12	TN	657	17
26	P	F	MS	TPR	Nicotine	5, 6, 7, …, 12	TN	234	13
27	H	F	MS	PU	Dark Green	5, 6, 7, …, 12	UG	329	13
28	Z	F	MS	PU	Dark Green	5, 6, 7, …, 12	UG	574	15
29	L	F	MS	PU	Honey	5, 6, 7, …, 12	UH	116	10
30	J	F	MS	PU	Nicotine	5, 6, 7, …, 12	UH	432	14
31	V	F	MS	TPR	Honey	5, 6, 7, …, 12	TH	354	13
32	R	F	MS	PU	Nicotine	5, 6, 7, …, 12	UN	230	11

Mathematical Model

The objective function is to minimize n_T and it is given in Equation (1). Each job can be processed only once as shown in Equation (2). Equation (3) shows that each position in each cell can be assigned to at most one job. Equation (4) enforces jobs to be assigned consecutively in each cell. Equation (5) controls setup requirements. If the consecutive jobs are from different families, then this constraint adds setup between those consecutive jobs. In Equations (6), (7.a) and (7.b), If-Then constraints are used to

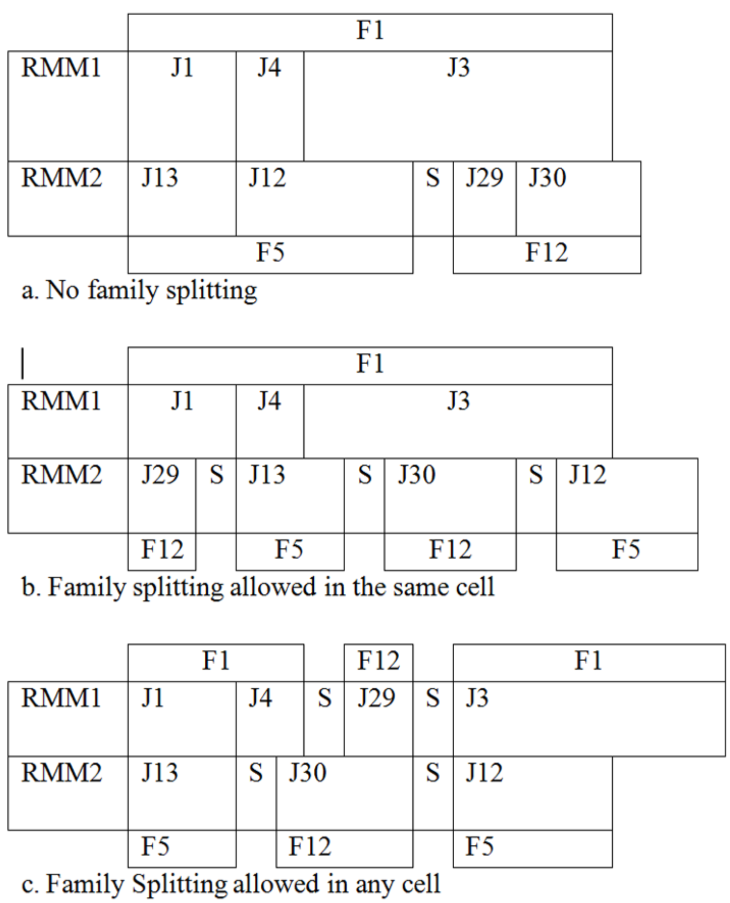
Table 2. Families for orders given in Table 1

Family No.	Jobs (due dates)
1	1 (10), 3 (22), 4 (11)
2	2 (12), 8 (15), 9 (11)
3	5 (20), 6 (17), 7 (13), 14 (14), 15 (10)
4	10 (14), 11(16), 16 (20)
5	12 (17), 13 (11)
6	17 (16), 18 (12)
7	19 (14), 20 (13), 21 (16)
8	22 (11), 31 (13)
9	23 (16), 32 (11)
10	24 (12), 25 (17), 26 (13)
11	27 (13), 28 (15)
12	29 (10), 30 (14)

Table 3. The processing times for examples of possible minicell loading cases

Family No.	Job No.	Processing Times (hrs)
F1	J1	3
	J3	9
	J4	2
F5	J12	5
	J13	3
F12	J29	2
	J30	4

Figure 5. Examples of possible minicell loading cases



eliminate the nonlinearity in the model. Equation (6) checks if a position is occupied by a job. If so, Equations (7.a) and (7.b) calculate the completion time of the job in that position. Equation (8) determines the tardiness value of a job. Equation (9) identifies if a job is tardy.

Notation

Indices:

- i Family index
- j Job index
- k Position index
- m Cell index

Parameters:

- n Number of jobs
- n_i Number of jobs in family *i*
- f Number of families
- M Number of cells
- P_{ij} Process time of job *j* from family *i*
- D_{ij} Due date of job *j* from family *i*
- S Setup Time
- R Very big number (larger than maximum possible tardiness value)

Decision Variables:

- Y_{mk} 0 if k^{th} position in cell *m* is occupied, 1 otherwise.
- X_{ijmk} 1 if job *j* from family *i* is assigned to the k^{th} position in cell *m*, 0 otherwise.
- C_{mk} Completion time of the job in k^{th} position in cell *m*
- T_{mk} Tardiness value of the job in k^{th} position in cell *m*
- W_{mk} 1 if setup is needed before the job in k^{th} position in cell *m*, 0 otherwise.
- nT_{mk} Coefficient for determining the tardiness of the job in k^{th} position in cell *m*. 1 if the job which is assigned to the k^{th} position in cell *m* is tardy, 0 otherwise.

Objective Function:

$$\min Z = \sum_{m=1}^{M}\sum_{k=1}^{n} nT_{mk} \quad (1)$$

Subject to:

$$\sum_{m=1}^{M}\sum_{k=1}^{n} x_{ijmk} = 1 \text{ for } i=1,\ldots,f,\ j=1,\ldots n_i \quad (2)$$

$$\sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijmk} \le 1 \text{ for } m=1,\ldots,M,\ k=1,\ldots,n \quad (3)$$

$$\sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijmk} \ge \sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijm(k+1)} \text{ for } m=1,\ldots,M,\ k=1,\ldots,n-1 \quad (4)$$

$$1 + W_{mk} \ge \sum_{j=1}^{n_i} x_{ijmk} + \sum_{j=1}^{n_q}\sum_{q\in(f/i)} x_{ijm(k-1)} \text{ for } m=1,\ldots,M,\ k=2,\ldots,n \quad (5)$$

$$\sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijmk} \le R * (1 - Y_{mk}) \text{ for } m=1,\ldots,M,\ k=1,\ldots,n \quad (6)$$

$$-C_{m1} + \sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijm1} * P_{ij} \le R * Y_{m1}\ m=1,\ldots,M \quad (7a)$$

$$C_{m(k-1)} - C_{mk} + S * W_{mk} + \sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijmk} * P_{ij} \le R * Y_{mk} \text{ for } m=1,\ldots,M,\ k=2,\ldots,n \quad (7b)$$

$$C_{mk} - \sum_{i=1}^{f}\sum_{j=1}^{n_i} x_{ijmk} * D_{ij} \le T_{mk} \quad (8)$$

for $m=1,\ldots,M,\ k=1,\ldots,n$

$$T_{mk} \le R^* nT_{nk};\ \text{for } m=1,\ldots,M,\ k=1,\ldots,n \quad (9)$$

Definition of Variables:

$x_{ijmk}\in\{0,1\}$; for $i=1,\ldots,f$, $j=1,\ldots,n_i$, $m=1,\ldots,M$, $k=1,\ldots,n$

$W_{mk}\in\{0,1\}$; for $m=1,\ldots,M$, $k=1,\ldots,n$

$nT_{mk} \in \{0,1\}$; for $m=1,\ldots,M$, $k=1,\ldots,n$

$Y_{mk} \in \{0,1\}$; for $m=1,\ldots,M$, $k=1,\ldots,n$

$C_{mk} \geq 0$; for $m=1,\ldots,M$, $k=1,\ldots,n$

$T_{mk} \geq 0$; for $m=1,\ldots,M$, $k=1,\ldots,n$

Genetic Algorithm

First, the initial population of n chromosomes is formed randomly. Then, mating partners are determined using mating strategies to perform crossover. The crossover and mutation operators are performed to generate offspring. For selecting the next generation, parents are added to the selection pool along with offspring. The next generation is selected from this pool based on their fitness function value. These steps are repeated until the number of the generations specified by the user is reached. Finally, the best chromosome obtained during the entire generations is determined as the final solution.

Chromosome Representation

The chromosome representation is used as an individual solution including genes corresponding to jobs. For each gene, the following representation code is used: (X, Y) where X denotes the Job Number and Y denotes the cell to which Job (X) is assigned. The sequence of genes in a chromosome also determines the sequence of jobs in the cells. As an illustration, an example is shown in Figure 6 where Jobs 1 & 3 are assigned to the first cell and Jobs 4 & 2 are assigned to the second cell in the order stated.

Figure 6. A Chromosome representation for a 4-job and 2-cell problem

1, 1	4, 2	3, 1	2, 2

Mating

Three different mating strategies are used to determine mating pairs; Random (R), Best-Best (B-B), and Best-Worst (B-W). The reproduction probabilities of the chromosomes are calculated according to their fitness function. The next step depends on the selected mating strategy.

If the Random Mating Strategy is selected, the mating pairs are determined randomly with respect to their reproduction probabilities by using Roulette Wheel approach. In this mating strategy, each chromosome and its randomly determined partner give one offspring. If the Best-Best Mating Strategy is selected, all chromosomes are ranked with respect to their reproduction probabilities in descending order. Then, the best chromosome is paired with the second best chromosome, the third chromosome paired with the fourth chromosome and so on. In addition, the first X% of the pairs produce 3 offspring, the next Y% of the pairs give 2 offspring, and the last Z% of the pairs produce 1 offspring. If the Best-Worst Mating Strategy is selected, all chromosomes are ranked with respect to their reproduction probabilities in descending order. Then, the best chromosome is paired with the worst chromosome; the second best chromosome is paired with the second worst chromosome and so on.

Crossover

Two different strategies are used to perform the crossover operation. Those crossover strategies are Position-Based Crossover (P-B) (Syswerda, 1999) and Order Crossover (OX) Strategies (Davis, 1985). The crossover operation is applied to the identified pairs with a probability of P_C. The first parent is copied as the offspring if crossover is not performed. The crossover operator affects the sequence of jobs but not their cell assignment.

In other words, the crossover is applied to the genes' X element and not to Y element.

Mutation

Two steps are used in the mutation operator. The first one is used for job sequence and only Reciprocal Exchange (R-E) Mutation Strategy is used [(see Gen and Cheng (1997)]. The mutation for job sequence is performed with a probability of P_{MJ}. The second step involves mutating cell assignments. In the mutation of cell assignments, two different mutation strategies are used, Random (R), and Reciprocal Exchange Mutation. The mutation of the cell assignment is performed with a probability of P_{MC}.

Selection

In this study, selection pool consists of all offspring and some of the parents. The next generation is selected from this pool. The selection from parents is a two-step process. First, the parents are ranked with respect to their reproduction probability in descending order. Then, the best P_E% parents are directly selected to advance to the selection pool. Then, the remaining (100-PE)% chromosomes are selected from the parents randomly based on their reproduction probability using Roulette Wheel Selection.

Once the selection pool is identified, the chromosomes are ranked with respect to their reproduction probability and a final selection is made from this pool to generate the next generation. In some experiments, we also allowed a certain percentage of lowest performers (Pw%) to advance automatically to the next generation to avoid immature convergence of the population.

ANALYSIS OF RESULTS

The results are grouped in four sections. 1) Genetic Algorithm Application, 2) Comparison of Mathematical Models with Genetic Algorithm, 3) Due Date Sensitivity Analysis, and 4) Setup Time Sensitivity Analysis. The experimental conditions are mentioned first and then the results obtained are discussed.

Data Sets Used

Nine data sets are used in the experimentation. The details of data sets are listed in Table 4. The data sets 1, 2, and 3 are realistic data sets obtained directly from the shoe manufacturing company. The data sets 4, 5, and 6 are relatively smaller data sets and they are generated from data sets 1, 2, and 3, respectively for only one cell. Similarly, for multiple cells; smaller data sets 7, 8, and 9, are generated from data sets 1, 2, and 3, respectively, by reducing batch sizes and due dates.

Genetic Algorithm Application

In this experiment, data sets 1, 2, and 3 are used. Initially, default parameters that are given in Table 5 are used. Then, the values of the GA parameters are changed one at a time in order to obtain better combinations. These parameters are listed in Table 6. Ten replications are performed.

The best solution is 3 tardy jobs for all three data sets. The frequencies of the best solution as well as the average values for better combinations are given in Tables 7, 8 and 9. For data set 1, four combinations stood out as the best combinations. Combination 1 has the highest frequency of the best solution. However, there is no significant difference among them according to ANOVA test results ($P=0.292 > \alpha$-value=0.05). For data set 2, top six combinations are determined as given in Table 8. The combination 3 has the best frequency among all combinations. Since there is significant difference between six combinations ($P=0.008<alpha=0.05$), Fisher Test is applied. The results show that combination 5 is significantly different (worst) than other five combinations. For data set 3, top six combinations are listed in

Table 4. Details of data sets

Data Set	Number of Families	Number of Jobs	Total Processing Time (minutes)	Number of Cells	Available Production Capacity (minute/cell)
Data 1	12	41	6465	3	2400
Data 2	12	33	6603	3	2400
Data 3	12	44	7995	4	2400
Data 4	5	11	2015	1	2400
Data 5	9	12	2146	1	2400
Data 6	9	13	2077	1	2400
Data 7	6	20	3222	3	1200
Data 8	5	13	2283	3	800
Data 9	6	22	4060	4	1200

Table 5. The default parameters for GA

Setup Time:	20 minutes
Population Size:	1000
Number of Generations:	1000
Elite Ratio:	0.2
Worst Ratio:	0.1
Crossover Probability:	1
Jobs Mutation Probability:	0.1
Cells Mutation Probability:	0.7
Mating Strategy:	Random
Crossover Strategy:	Position-Based
Mutation Strategy:	Random

Table 6. Values of the parameters for GA

PARAMETER	VALUES			
Elite Ratio:	0.2	0.1	0.3	
Worst Ratio:	0.1	0	0.2	0.3
Crossover Probability:	1	0.7	0.5	
Jobs Mutation Probability:	0.10	0.05	0.2	
Cells Mutation Probability:	0.7	0.5	0.3	
Mating Strategy:	Random	Best-Best	Best-Worst	
Crossover Strategy:	Position-Based	Order		
Mutation Strategy:	Random	Exchange		

Table 9. Combination 2 has the largest frequency for the best solution. Since there is significant difference (P=0.034<α-value=0.05) between six top combinations, Fisher Test is applied. The results show that Combinations 1 and 2 are significantly different than Combinations 5 and 6. Combination 3 and Combination 4 are also eliminated as they are not significantly different than Combinations 5 and 6.

Comparison of Mathematical Models and Genetic Algorithms

In this section, the results of the Mathematical Models are compared with the GA results for the modified data sets. The experimental conditions that are detailed in 1 are utilized in GA application. The Mathematical Model solutions (optimal solutions) and GA solutions for data sets 4, 5, and 6 are given in Table 10; for data sets 7, 8, and 9 are given in Table 11.

GA found the optimal solution ten times out of ten replications for both one cell and multiple cells (except data set 7). For data sets 4, 5, and 6, Mathematical Model has a better execution time compared to GA (except data set 5) due to small problem size. When data sets 7, 8, and 9 are considered, the GA has significantly better execution time compared to the Mathematical Model. This is expected since as the problem size increases; the execution time of the Mathematical Model increases dramatically. The details of data set 8 and corresponding Gantt chart for the mathematical model results are given in Appendix for illustration purposes.

Due Date Sensitivity Analysis

For due date sensitivity analysis, data set 1 is used. Combination 4 is selected as the parameters of GA to perform this experiment. In this experiment, due date sets of loose (Loose 1), looser (Loose 2),

Table 7. The best combinations for data set 1

	Elite Ratio	Worst Ratio	Crossover Prob.	Jobs Mut. Prob.	Cells Mut. Prob.	Mating Strategy	Crossover Strategy	Mutation Strategy	Freq of Best Known Sol	Average nT
Comb. 1	0.3	0	1	0.05	0.7	B-W	P-B	R	4	3.7
Comb. 2	0.3	0	1	0.05	0.7	B-B	P-B	R	4	3.6
Comb. 3	0.3	0	1	0.05	0.7	B-W	OX	R	1	4.1
Comb. 4	0.3	0	1	0.05	0.7	B-B	OX	R	3	3.8

Table 8. The best combinations for data set 2

	Elite Ratio	Worst Ratio	Crossover Prob.	Jobs Mut. Prob.	Cells Mut. Prob.	Mating Strategy	Crossover Strategy	Mutation Strategy	Freq of Best Known Sol	Average nT
Comb. 1	0.1	0	0.7	0.05	0.5	R	P-B	R	7	3.3
Comb. 2	0.1	0	0.7	0.05	0.5	B-B	P-B	R	7	3.3
Comb. 3	0.1	0	0.7	0.05	0.5	B-W	P-B	R	**8**	**3.2**
Comb. 4	0.1	0	0.7	0.05	0.5	R	OX	R	6	3.4
Comb. 5	0.1	0	0.7	0.05	0.5	B-B	OX	R	1	4
Comb. 6	0.1	0	0.7	0.05	0.5	B-W	OX	R	5	3.5

Table 9. The best combinations for data set 3

	Elite Ratio	Worst Ratio	Crossover Prob.	Jobs Mut. Prob.	Cells Mut. Prob.	Mating Strategy	Crossover Strategy	Mutation Strategy	Frequency of Best Known Solution	Average n_T
Comb. 1	0.3	0	0.7	0.05	0.5	R	P-B	R	4	3.7
Comb. 2	0.3	0	0.7	0.05	0.5	B-B	P-B	R	**5**	**3.7**
Comb. 3	0.3	0	0.7	0.05	0.5	B-W	P-B	R	1	4.2
Comb. 4	0.3	0	0.7	0.05	0.5	R	OX	R	0	4.2
Comb. 5	0.3	0	0.7	0.05	0.5	B-B	OX	R	0	4.4
Comb. 6	0.3	0	0.7	0.05	0.5	B-W	OX	R	0	4.5

Table 10. The optimal solutions and GA results for minimizing nT for one cell

Data Set	Math Model Result (Opt. Sol.)	Decision Variable	Constraints	Math Model Execution Time (hr:min:sec)	Optimal Frequency for GA (x/10)	GA Execution Time (hr:min:sec)
Data 4	2	166	106	00:00:02	10	00:00:53
Data 5	3	193	160	00:04:56	10	00:00:57
Data 6	1	222	174	00:00:16	10	00:01:04

Table 11. The optimal solutions and GA results for minimizing nT for multiple cells

Data Set	Math Model Result (Opt. Sol.)	Decision Variable	Constraints	Math Model Execution Time (hr:min:sec)	Optimal Frequency for GA (x/10)	GA Execution Time (hr:min:sec)	Average of GA Results (n_T)
Data 7	2	1499	720	56:01:00	2	00:01:52	2.8
Data 8	4	701	425	16:21:17	10	00:01:03	4
Data 9	2	2374	1051	05:54:40	10	00:01:49	2

tight (Tight 1), and tighter (Tight 2) are generated in addition to the original due date set (Medium). The cumulative probabilities of those due date sets are given in Table 12. The results of ten replications for the due date sets are given in Table 13. As expected, the number of tardy jobs increased as due dates got reduced.

Setup Time Sensitivity Analysis

The set up times varied from 0 to 100 for data set 1. Similarly, Combination 4 parameters were used in this experiment. The results of GA for the various setup times are given in Table 14. The number of tardy jobs increased as the setup time increased. This was expected, however, in some cases setup times doubled (from 5 minutes to 10 minutes; from 10 min to 20 min; from 20 min to 40 min) but number of tardy jobs increased only by one. On the other hand, when setup time increased from 40 minutes to 80 minutes, the number of tardy jobs increased by three. This shows that system can tolerate increase in setup times to a certain extent, beyond which its impact will be bigger.

Table 12. The cumulative probabilities of different due date sets

Due Date (min.)	480	960	1440	1920	2400
Loose 2	0	0	0.33	0.67	1
Loose 1	0	0.25	0.5	0.75	1
Medium	0.2	0.4	0.6	0.8	1
Tight 1	0.25	0.5	0.75	1	0
Tight 2	0.33	0.67	1	0	0

Table 13. The nT results of GA for different due date sets

	DUE DATE				
Replications	**Loose 2**	**Loose 1**	**Medium**	**Tight 1**	**Tight 2**
1	0	0	3	6	13
2	0	0	4	7	14
3	0	0	4	8	13
4	0	0	3	6	13
5	0	0	3	6	14
6	0	0	4	7	13
7	0	1	4	7	13
8	0	0	5	7	13
9	0	0	5	7	13
10	0	0	3	6	13

Table 14. The nT results of GA for different setup times

Replications	Setup Time (min.)							
	0	**5**	**10**	**20**	**40**	**60**	**80**	**100**
1	1	3	3	3	6	8	8	10
2	2	3	3	4	6	7	8	11
3	2	2	3	4	7	9	9	11
4	2	3	2	3	5	8	10	10
5	1	2	3	3	5	8	9	9
6	2	3	3	4	5	8	9	11
7	2	3	3	4	5	7	9	10
8	2	3	4	5	6	7	10	10
9	1	2	3	5	7	7	9	10
10	3	2	2	3	6	8	10	10

CONCLUSION AND FUTURE WORK

In the problem studied in this chapter, every job has individual due dates even the ones in the same family. This property of the problem completely separates this study from other cellular manufacturing scheduling problems. The reason of this complexity is because of the natural conflict between meeting due dates of jobs and reducing total setup times. If the entire family is scheduled together, then the total setup time is at minimum. But, the jobs in the consecutive families may be forced to be postponed and probably become tardy. In contrast, splitting a family several times may increase the number of setups which reduce the productive time, and finally have an adverse effect on the number of tardy jobs.

The Mathematical Model is one of the solution techniques which guarantee to find the optimal solution. It is not practical to solve larger problems using the mathematical models because of the complexity. As a result, there is a need to use other approaches to solve such problems. We proposed and used Genetic Algorithm approach in this Chapter.

Genetic Algorithms found the optimal solution in all problems with varying frequency. The execution time of Mathematical Model was reasonable only for small problem sizes. GA clearly outperformed Mathematical Model with respect to execution times.

The results showed that family splitting occurred in all multi-cell problems. Due to limited space, we presented only one in this chapter. The occurrence of family splitting in these problems show us that the system used the feature of family splitting since it was beneficial in terms of reducing the number of tardy jobs. Another conclusion that can be drawn is that the impact of setup times and due dates on the system performance was as expected.

We are planning to extend this work to sequence-dependent setup times in the future and also use other metata-heuristic techniques. This work can also be extended to include other performance measures and job-splitting option.

REFERENCES

Askin, R. M., & Standridge, C. R. (1993). *Modeling and analysis of manufacturing systems* (1st ed.). New York, NY: Wiley.

Davis, L. (1985). Applying adaptive algorithms to domains. In *Proceedings of the International Joint Conference on Artificial Intelligence*, (pp. 162-164).

Gen, M., & Cheng, R. (1997). *Genetic algorithms and engineering design*. New York, NY: Wiley.

Gupta, J. N. D., & Chantaravarapan, S. (2008). Single machine group scheduling with family setups to minimize total tardiness. *International Journal of Production Research*, *46*(6), 1707–1722. doi:10.1080/00207540601009976

Ham, I., Hitomi, K., Nakamura, N., & Yoshida, T. (1979). Optimal group scheduling and machining-speed decision under due-date constraints. *Transactions of the ASME: Journal of Engineering for Industry*, *101*, 128–134. doi:10.1115/1.3439485

Hitomi, K., & Ham, I. (1978). Machine loading and product-mix analysis for group technology. *Transactions of the ASME: Journal of Engineering for Industry*, *100*, 370–374. doi:10.1115/1.3439443

Nakamura, N., Yoshida, T., & Hitomi, K. (1978). Group production scheduling for minimum total tardiness. *AIIE Transactions*, *10*(2), 157–162.

Pan, J. C., & Wu, C. (1998). Single machine group scheduling to minimize mean flow time subject to due date constraints. *Production Planning and Control*, *9*(4), 366–370. doi:10.1080/095372898234091

Subramanian, A. K. (2004). *Cell loading and scheduling in a shoe manufacturing company*. Unpublished Master's Thesis, Ohio University, Athens.

Süer, G. A., Arikan, F., & Babayigit, C. (2008). Bi-objective cell loading problem with non-zero setup times with fuzzy aspiration levels in labour-intensive manufacturing cells. *International Journal of Production Research*, *46*(2), 371–404. doi:10.1080/00207540601138460

Süer, G. A., Arikan, F., & Babayigit, C. (2009). Effects of different fuzzy operators on fuzzy bi-objective cell loading problem in labor-intensive manufacturing cells. *Computers & Industrial Engineering*, *56*, 476–488. doi:10.1016/j.cie.2008.02.001

Süer, G. A., Cosner, J., & Patten, A. (2009). Models for cell loading and product sequencing in labor-intensive cells. *Computers & Industrial Engineering*, *56*, 97–105. doi:10.1016/j.cie.2008.04.002

Suer, G. A., & Dagli, C. (2005). Intra-cell transfers and cell loading in labor-intensive manufacturing cells. *Computers & Industrial Engineering*, *48*, 643–655. doi:10.1016/j.cie.2003.03.006

Süer, G. A., Saiz, M., Dagli, C., & Gonzalez, W. (1995). Manufacturing cell loading rules and algorithms for connected cells. In Kamrani, A., Parsaei, H., & Liles, D. (Eds.), *Planning, design, and analysis of cellular manufacturing systems* (pp. 97–127). New York, NY: Elsevier. doi:10.1016/S1572-4417(06)80038-0

Süer, G. A., Saiz, M., & Gonzalez, W. (1999a). Evaluation of manufacturing cell loading rules for independent cells. *International Journal of Production Research*, *37*(15), 3445–3468. doi:10.1080/002075499190130

Süer, G. A., Santos, J., & Vazquez, R. (1999b). Scheduling rotary injection molding machines. *Proceedings of the Second Asia-Pacific Conference on Industrial Engineering and Management Systems,* Kanazawa, Japan, (pp. 319-322).

Süer, G. A., Subramanian, A., & Huang, J. (2009). Heuristic procedures and mathematical models for cell loading and scheduling in a shoe manufacturing company. *Computers & Industrial Engineering*, *56*, 462–475. doi:10.1016/j.cie.2008.10.008

Süer, G. A., Vazquez, R., & Cortes, M. (2005). A hybrid approach of genetic algorithms and local optimizers in cell loading. *Computers & Industrial Engineering*, *48*, 625–641. doi:10.1016/j.cie.2003.03.005

Suresh, N. C., & Kay, J. M. (1998). Group technology & cellular manufacturing: Updated perspectives. In Suresh, N., & Kay, J. (Eds.), *Group technology and cellular manufacturing* (pp. 1–14). Kluwer Academic Publishers. doi:10.1007/978-1-4615-5467-7_1

Syswerda, G. (1991). Scheduling optimization using genetic algorithms. In Davis, L. D. (Ed.), *Handbook of genetic algorithms* (pp. 332–349). New York, NY: Van Nostrand Reinhold.

Urs, B. S. (2005). *Comparison of heuristic procedures for scheduling rotary injection molding machines*. Unpublished Master's Thesis, Ohio University, Athens.

Yarimoglu, F. (2009). *Cell loading and product sequencing subject to manpower restrictions in synchronized manufacturing cells*. Unpublished Master's Thesis, Ohio University, Athens.

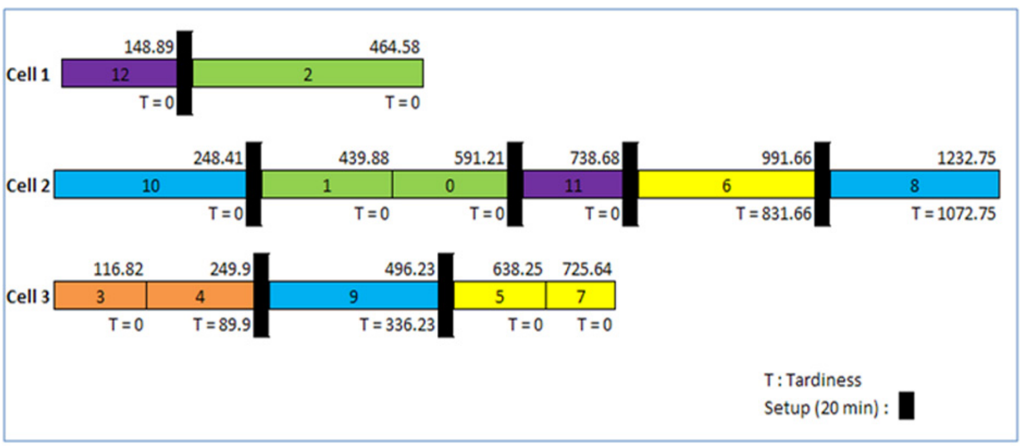
APPENDIX

Optimal Solution for Data Set 8

Table 15. Data Set 8

Job Number	Family No	Processing Time	Due Date
0	PB	151.324	800
1	PB	171.471	800
2	PB	295.684	480
3	PG	116.815	160
4	PG	133.081	160
5	PH	122.016	640
6	PH	232.978	160
7	PH	87.392	800
8	PN	221.099	160
9	PN	226.336	160
10	PN	248.412	320
11	TB	127.47	800
12	TB	148.891	160

Figure 7. Gantt chart for the optimal solution for data set 8



Chapter 65
Evaluation of Key Metrics for Performance Measurement of a Lean Deployment Effort

Edem G. Tetteh
Paine College, USA

Ephrem Eyob
Virginia State University, USA

Yao Amewokunu
Virginia State University, USA

ABSTRACT

To meet customer's needs for high-quality goods and avoiding risks of product-liability, global firms continually evaluate the performance of their supply chain for optimum design. Lean management is one of the key techniques businesses adopt in redesigning their processes. The technique is a vital strategy to increase productivity and effectiveness with respect to the movement of goods. Multivariate Analysis of Variance (MANOVA) was utilized to evaluate the performance of work cell, shift, worker's experience, and kaizen event participation level during a lean enterprise deployment effort at a multinational organization. The significance of the effects of these variables were assessed based on various lean supply chain factors such as First In First Out (FIFO), Setup Wheel System (SWS), Standard Operating Procedures (SOP), Clip System(CS), and Key Performances Indicators (KPI). The results support the criticality of metrics and their impact in implementing a lean manufacturing process in a global supply chain environment.

INTRODUCTION

Lean manufacturing has been shown to improve the competitiveness of organizations. The concept of lean started first with the Toyota Production System (TPS). Then, the idea was expanded by a research group at Massachusetts of Institute of Technology (Womack & Jones, 1996). The philosophy of lean uses a process of waste reduction, thus producing higher quantity and better quality products with the least resources possible. The

DOI: 10.4018/978-1-4666-1945-6.ch065

goals are zero lead time, zero inventory, and zero defects, resulting in higher customer satisfaction (Tapping, Luyster, & Shuker, 2002).

Lean manufacturing strives for continuous improvement towards an ideal through relentless reduction of waste, where ideal means delivering what the customer requests on time, on demand, and free of defects (Miller, 2005). Lean process can be traced back to the early 1900s, when Henry Ford introduced the notion of mass production in 1913 (Miller, 2005). However, according to Soderquist and Motwani (1999), Taiichi Ohno was the first to present lean manufacturing to eliminate production waste at Toyota (Soderquist & Motwani, 1999). Source of wastes identified by Ohno include: errors that require recertification, product defects, process steps that are not needed, goods or employee movement without any purpose, goods and services that don't meet the needs of the customers, and any waiting time due to bottlenecks. Taiichi Ohno classified these wastes into seven basic types: overproduction, transportation, process waste, operator movement, inventory, idle time, bad quality (Bateman & David, 2002). Continuous improvement with a focus on the seven wastes is part of lean manufacturing.

It uses a team focused methodology requiring knowledge to be pulled from everyone (from the hourly worker to the upper management), and driven by continuous improvement (Kaizen). The team focuses on continuous improvements and uses tools and techniques to identify and eliminate wastes. Harris and Donatelli (2005) stated that value-stream mapping is the foundational tool used by any company that is on the cutting edge of transformation from a traditional organization to a lean enterprise. Value-stream mapping of products was started through the Toyota production system of lean manufacturing. Ninety-four percent of manufacturing errors or problems belong to systems, and lean manufacturing attacks these systems with the common goal of cost reduction or improvement of production (Deming, 1986).

Kaizen events are ways of accelerating improvements to worker productivity. These events help management to find new ways to gain substantial savings in time, space and labor output (Alukal & Manos, 2006). During Kaizen events, worker's ideas are highly encouraged for frequent and small improvements. This results in shrunken lead times, dramatic reduction in work-in-process, and reduction of scrap and defects, while minimizing the need for capital expenditures (Mika, 2005). They are important because they provide an excellent return on investments of financial and human resources. Furthermore, continued improvements will compound the return, since Kaizen never really ends (Mika, 2005). The events often eliminate the need for costly overtime by improving processes while collapsing lead times, and dramatically reducing work-in-process. It helps focus on improving material flow, information flow, and process quality of a business.

The rest of the chapter will cover the following sections: First, a description of lean techniques is covered. Next, various lean enterprise tools are discussed. Finally, a case study of a firm is presented. The research method, data collection, analyzes, and results are covered.

LEAN TECHNIQUES

One of the primary lean techniques is Poke Yoke. It was developed by Shingo Shigeo in the 1960's, and can be seen as the art of error proofing (Elbadawi, McWilliams and Tetteh, 2010). The process is designed to make it hard to make mistakes or at least easy to be detected and corrected. According to the authors, Shingo explained that a mistake is something a human cannot avoid, but a defect is allowing such a mistake to reach the customer. Therefore, defects are totally avoidable. Poke Yoke utilizes set-up devices or inspection techniques to ensure that the process is done correctly. Standardized works set a foundation that facilitates future change and continuous improvement (Zimmerman, 2010). In many organizations, work procedures called standardized work are written with highly detailed descriptions, but still holding their simplicity (Emiliani, 2008). These work procedures remove unneeded variation and confusion, allowing the process to flow smoothly.

Another technique used by companies to tackle waste is quick changeover. It's the act of changing from the last good piece of a product run to the first good piece of the next run. Quick changeover minimizes the time that operators use during changeover operations, since the excess time doesn't add any value to the products or process. The goal is to spend less time changing from one job to the other. This allows new employees to comprehend complicated operations easily and cover other operations that are not their specialization (Shingo, 1985). The technique helps achieve a smooth flow of value on a production line of different products. One more lean technique that goes along with quick changeover is 5S. It's a series of lean manufacturing standards used to improve the work place (Cross, 2008). 5S are cleansing methods that provide workers with a comfortable and efficient workspace enabling them to perform at their most effective level.

Furthermore, Total Productive Maintenance (TPM), another lean technique, was introduced by Seiichi Nakajima in the late 1970's (Elbadawi, McWilliams & Tetteh, 2010). It aims to ensure the effectiveness of production equipment. TPM focuses on ridding time loss due to breakdowns and stoppages. The origin of TPM is Nipponese, a Japanese manufacturer of automotive electrical parts. The method exerted a major influence over the economic progress made by Japanese manufacturers in the late 1970's. It's based on a continuous improvement strategy that embraces all aspects of the organization maximizing the productivity of the equipment. It also contributes to a successful strategy towards zero breakdowns, zero defect, and lower cost, which can be achieved by observing the equipment's life span. The most popular performance measure in TPM is the Overall Equipment Effectiveness (OEE) defined as follows (Smith & Hawkins, 2004):

OEE= Equipment Availability × Performance Efficiency × Rate of Quality

The world-class level of OEE starts at 85% based on the following values: 90% equipment availability, 95% performance efficiency, and 99% quality rate (Evan & Lindsay, 2004). This is achievable in concordance with Total Quality Management (TQM). Evan & Lindsay (2004) described TQM as the integration of all functions and processes within an organization to achieve a continuous improvement in the quality of goods and services. The technique is used to improve continuously all areas of a company's operation to satisfy customers' requirements (Liang, 2010).

Kanban is a key element in a just-in-time system (Widyadana, 2010). It's a signaling system that gives instructions on when to manufacture or supply a component based on the actual usage of the material. It also informs on what, when and how much should be produced. However, the technique can cause dramatic spikes in demand with the influx of new orders. Kaizen events can be used to minimize the disruption. It's a technique that focuses on structured improvement projects, using a dedicated cross-functional team to improve a targeted work area, with specific goals, in an accelerated timeframe (Farris et al., 2009). Six Sigma methodology and Lean management tools are utilized to find the root causes of issues in many areas of an organization. This provides organizations with management tools needed to make their processes more efficient and effective to serve customers better.

LEAN ENTERPRISE TOOLS

First In First Out (FIFO) is a useful lean tool when dealing with logistics and inventory management. It facilitates order sequencing and control approach

Figure 1. FIFO sign with 18 reels maximum

in fulfilling demands. It prevents earlier orders from being delayed in favor of newer commands. FIFO is used to help operators to complete orders in the sequence that they are received, and up to the maximum. An example of an application of FIFO is displayed in Figure 1. This allows the company to ensure that customer's orders received first are completed and shipped first.

For a maximum efficiency of FIFO, a customized tool namely setup wheel was designed. As illustrated in Figure 2, this tool was based on a set of racks used to stop the reels from moving freely. The tool facilitates the organization of work area by physically straightening up the reels to reflect the sequences in the orders. Prior to use of this tool, it was difficult for operators to prevent the reels to roll or move to undesired locations.

Standard Operating Procedures (SOPs) are written instructions that outline the steps or tasks needed to complete a task operate a piece of machinery or plant. SOPs are effective parts of a quality system with a purpose to assist workers to carry out operations correctly and consistently. The tool enables the workers in the study plant to understand the steps in performing their task. As shown on Form 1, the instructions were written in a brief manner with simple language that all operators are able to read and understand.

One additional tool that helped the implementation of the FIFO is the clip system. It consists of customized color coded clips designed to signify the maximization of the FIFO lanes when they are at full capacity. One of its main goals is to prevent the overflow of the setup wheel system, and help the management of the orders. The colors were used to identify bathes. For example, the color red can be used to identify the first order of twenty reels on a thirty reels maximum FIFO lane and the yellow color will be used for a second set of ten reels on the same lane. Other basic customer information such as customer's names, locations, order and cell numbers, and operator's identification were printed on the clip.

All the lean tools discussed in the previous paragraphs constituted the key performance indicator (KPI). In lean manufacture, KPI is a useful tool for measuring and developing metrics that support and facilitate the achievement of the critical goals of the organization (Vorne, 2011). Key performance indicators are important for understanding and improving manufacturing performance both from lean manufacturing perspective of eliminating waste and from corporate point of view of achieving strategic goals. The tool can be highly effective for exposing, quantifying and visualizing muda, the lean term for waste (Vorne, 2011). To be efficient, the measures must provide meaningful, reliable, and accurate information. They are to be as specific as possible based on

Figure 2. A cell using a FIFO system

data collected. Companies' goals and desires are often vague, but KPIs are to craft them to be very specific.

STUDY METHOD

Study Settings

The study was conducted at a world-class manufacturing plant that makes signal transmission products primarily for the entertainment, residential, industrial and security markets. The plant is located in one of the Midwestern States of the United States, and it is home to one of the firm's global manufacturing divisions along with a distribution and an engineering center. At the site, workers design, manufacture, and market various product types such as cables, ethernet switches, and industrial connectivity for the following markets: industrial; audio and video; security; networking; and communications.

The company is one of the largest U.S. based manufacturers of high-speed electronic cables with specialty on electronics and data networking products. It has presence in the following four segments: Americas segment, Europe, Middle East, and Africa (EMEA) segment, Asia Pacific segment, and the Wireless segment.

The Americas segment contributed approximately 54%, 52%, and 57% of the company consolidated revenues consecutively in 2009, 2008, and 2007 respectively with an industrial market to include applications ranging from advanced industrial networking and robotics to traditional instrumentation and control systems. Their cable products are used in discrete manufacturing and process operations involving the connection of computers, programmable controllers, robots, operator interfaces, motor drives, sensors, printers, and other devices. They also cover other industrial environments, such as petrochemical and other harsh-environment operations, which require cables with exterior armor or jacketing to endure physical abuse, exposure to chemicals, extreme temperatures, and outside elements.

Procedure

The goal of the study was to assess performance metrics through a worker's level of satisfaction and learning after a 30 days deployment of a lean enterprise effort. During a month long intervention that preceded the survey, an audit form shown in Figure 3 was used to collect the work cells level of compliance with the FIFO, Setup Wheel, Standard Operating Procedures (SOP), Clip, and KPI systems.

The tasks consisted of the shift supervisors conducting a walk-through every hour to assess the following for compliance:

1. Is the number of reels exceeding the requirement on the FIFO sign?
2. Are the reels stored in an orderly manner on the tracks/lanes provided?
3. Are all colored clips accounted for? (All 10 clips for each color should either be on a reel or on a stack.)
4. Are any reels in the FIFO track/lane without an attached colored clip?

In the case of non-compliance, a shift supervisor would proceed to provide feedback to the operators in the work cell in question, providing them instructions on how to avoid a non-conformance in the future. The same evaluation course of action was repeated every hour during the three shifts for thirty days. Along with the short guidance provided by the shift supervisors for non-compliances, workers participated in many Kaizen events during the entire time of the study. At the conclusion of the 30 days intervention period, a learning and satisfaction assessment survey was given to the workers. One of the work cells in the plant, the "R cell", was used as a control group: workers assigned to that cell didn't participate in

Figure 3. FIFO audit form for the H cell

H2 FIFO MAX Audit

Auditor:________________________ Week:______________

	Item		1	2	3	4	5
PE 303 to Cabler 313 (Ceeco) FIFO MAX	1	Number of reels does not exceed FIFO MAX.					
	2	Reels are stored in an orderly manner on the tracks/lanes provided.					
PE 303 to Cabler 319 (Cook) FIFO MAX	3	Number of reels does not exceed FIFO MAX.					
	4	Reels are stored in an orderly manner on the tracks/lanes provided.					
Cabling to PE 310 FIFO Max	5	Number of reels does not exceed FIFO MAX.					
	6	Reels are stored in an orderly manner on the tracks/lanes provided.					
	7	All Colored Clips are accounted for. (All 10 Clips for each Color should either be on a reel or on a stack.)					
	8	No Reels should be in the FIFO MAX track/lane without a Colored Clip attached.					
OVERALL		**Total Possible Score**	10	10	10	10	10
		Score Out of Total Possible					
		Green (100%) or Red (<100%) Of Possible					

Next Page For Countermeasures

Item (1-10) Missed	Countermeasure	Responsible

any kaizen events, and received no feedback or instructions from their shift supervisors.

Analysis and Results

The independent variables (IV) were cell at four levels (C cells, H cell, M cells, and R cell); shift at three levels (1st, 2^{nd} and 3^{rd} shifts); worker's experience at five levels(Less than a year, 1-5 years, 6-15 years, 16-30 years, and more than 30 years); and worker's participation in Kaizen events at two levels (participation and non-participation. As illustrated in Table 1, the dependent variables (DV) were the work cell performance scores in the learning and satisfaction with lean tools such as FIFO, Setup Wheel, SOP, Clip System, and KPI.

SAS 9.1 software was used to analyze the data with the alpha level set at *0.05*. Tests were conducted to determine if there are statistical differences between and among the performance in work cell assignment, shift, experience, and worker's kaizen event participation based of the

Table 1. Description of the variables

Variable	N	Mean Score	Standard Deviation	Min	Max
FIFO	147	54.198318	40.283943	0	100
Setup Wheel	147	40.477961	40.115235	0	100
SOP	147	40.080056	38.956114	0	100
Clip System	147	42.507987	41.473942	0	100
KPI	147	36.457426	35.543168	0	100

cell performance on the following lean tools: FIFO, Setup wheel system, SOP, Clip System, and KPI.

A multivariate analysis of variance (MANOVA) was performed since there are five performance measures, some or all of which may be correlated.

The results showed that the MANOVA test is significant for *cell,* Λ=.72, $F(15, 339.95) = 2.89$, p <.001, and for *kaizen*, Λ=.86, $F(5, 123) = 3.98$, p =.002 as illustrated in Table 2. All the four cells are statistically different based on the variables FIFO, Setup wheel, SOP, Clip system, and KPI. This can be interpreted as people who participated in Kaizen events and those who did not significantly differ from each other.

On the basis of the previous results, another MANOVA was performed to test the main and interaction effects of variables *cell* and *kaizen.* They are found to be the only two dependent variables that significantly affect the set of variables under question 5. The results shown in Table 3 demonstrated that the interaction is not significant Λ=.93, $F(15, 353.75)$ =.62, p =.86. There were only main MANOVA effects of Kaizen and cell on the set of dependent variables under consideration. More specifically, cells means on FIFO, Setup wheel, SOP, Clip system, and KPI, as a set, are not equal, Λ=.75, $F(15, 353.75) = 2.63$, p <.001. Furthermore, the results show a significant association between participation in kaizen events and the same set of variables, Λ=.91, $F(5, 128) = 2.62$, p <.05. This can be interpreted as the performance of people who participate in Kaizen events differed significantly from employees who didn't take advantage of the training.

Since significant results were found in MANOVA, a Univariate analysis, namely an unbalanced two-way ANOVA, followed by multiple comparison procedure, was performed as a follow-up to verify which of the independent variable(s) is (are) associated with *cell* and *kaizen* respectively. The results show that the cells significantly differ on Clip system, $F(3, 132) = 5.17$, p <.01) whereas the differences between the cells are statistically non significant with regard to FIFO, Setup wheel, SOP, and KPI. The multiple

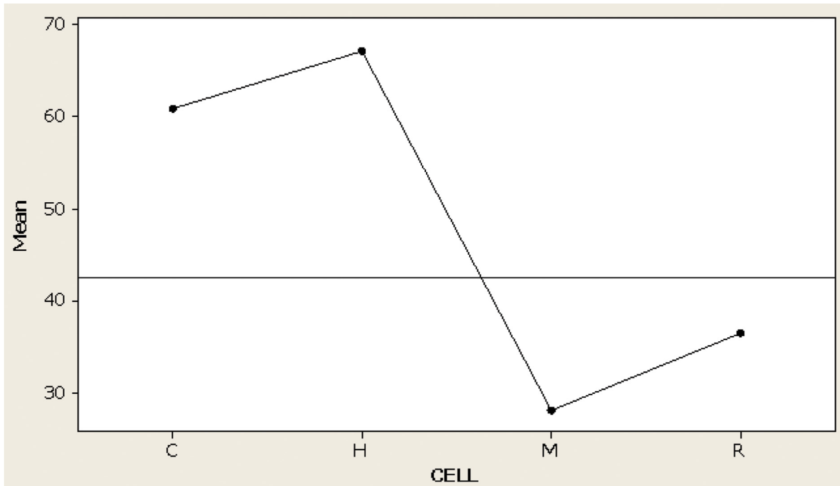
Table 2. MANOVA test for cell, shift, experience and kaizen effects on lean tools

	Λ	*F*	*dl*	*p*	η²
cell	.72	2.89	15, 339.95	<.001	.28
shift	.93	.91	10, 246	.525	.07
exp	.83	1.21	20, 408.89	.240	.17
kaizen	.86	3.98	5, 123	.002	.14

Table 3. MANOVA test for cell and kaizen effects on lean tools

	Λ	*F*	*dl*	*p*	η²
cell	.75	2.63	15, 353.75	<.001	.25
kaizen	.91	2.62	5, 128	.027	.09
Cell*kaizen	.93	.62	15, 353.75	.86	.07

Figure 4. Main effect for clip system



comparison analysis based on Dunnett procedure, reveals that cell R (the control group) differs from cell C and cell H but not from cell M. Contrary to what could be expected, theses groups know about lean tools better than the control group as illustrated in Figure 4.

Two models of MANOVA were tested. In the first model, Question 8 and Question 10 were included separately (or count as individual variables) in the set of the dependent variables. In the second model, the two variables were merged to create one variable called *cell performance*. To create *cell performance* we took the mean of Question 8 and Question 10 after Question 10 score was reversed. The results of these analyses are shown below in Table 4.

Table 4. MANOVA test for cell, and kaizen effects on lean tools

	Λ	*F*	*dl*	*p*	η²
cell	.75	2.63	15, 353.75	<.001	.25
kaizen	.91	2.62	5, 128	.027	.09
Cell*kaizen	.93	.62	15, 353.75	.86	.07

The result for the first model show that MANOVA test is significant, Λ=.85, $F(12, 365.41) = 1.94$, $p < .05$ as illustrated in Table 5; meaning that all the four cells have not equal means on the set of the variables comprising Question 7-Question 10.

As in the case of the lean tools, a univariate analysis of variance using GLM method was performed to identify which, if any, of the in dependent variables were associated with the variable *cell*. The univariate analysis results including multiple comparison show that the cells differ with regard to Question 7, $F(3, 141) = 4.18$, $p < .01$ and Q9 $F(3, 141) = 4.23$, $p < .01$.

The control group tends on average to be more satisfied by the way FIFO was implemented than the group in cell M. The two groups also differ with regard to Question 9, with cell R more likely to avoid confusion than cell M. Compared to the two remaining groups (cell C and cell H) the dif-

Table 5. MANOVA test for cell effect on Q7-Q10

	Λ	*F*	*dl*	*p*	η²
cell	.85	1.94	12, 365.41	.03	.15

Table 6. MANOVA test for cell effect on Q7, Q9 and cell performance

	Λ	*F*	*dl*	*p*	η²
cell	.88	2.11	9, 338.44	.03	.12

ferences are not statistically significant. These facts were expected because, by not having a direct participation in the lean effort at the plant, the workers in cell R were satisfied with the lean implementation effort. Cell R didn't have enough exposure and involvement to provide a good assessment of the FIFO system implementation.

The results of the second model are similar to the previous ones, except that cell performance is associated with cell. Thus MANOVA test is significant, Λ=.88, $F(9, 338.44) = 2.11$, $p < .05$ as shown in Table 6.

The univariate analysis of variance shows that each of the three independent variables is associated to variable *cell*. The results demonstrate that the means of the four cells regarding Question 7 are not all equal $F(3, 141) = 4.18$, $p < .01$and the same is true for Q9 $F(3, 141) = 4.23$, $p < .01$ and variable *cell performance* $F(3, 141) = 2.64$, $p = .05$.

Nonetheless, the multiple comparison results for the *cell performance* do not yield significant differences between the four groups. This is not a big surprise, since the exact *p* value for the omnibus test is.052, and the likelihood of finding significant differences through multiple comparisons was very slim. As for the results for Question 7 and Question 9, they remain unchanged.

CONCLUSION

The purpose of a lean system is to increase the throughput of an organization to meet customers' product demands. This will provide any firms with what they want, when they want it, and at the same time meets their customers' quality expectations. In a global supply chain environment, the identification of key performance measures and their assessment is critical to improve processes continuously. The results of this study showed that the reliance on systematic strategies using lean tools and principles as well as the emphasis on training are the key elements to a successful lean enterprise deployment in a global supply chain environment.

REFERENCES

Alukal, G., & Manos, A. (2006). *Lean kaizen: A simplified approach to process improvements*. Milwaukee, WI: American Society of Quality Press.

Bateman, N., & David, A. (2002). Process improvement programmes: A model for assessing sustainability. *International Journal of Operations & Production Management*, *22*(5), 515–526. doi:10.1108/01443570210425156

Condi, C. (2009). *Lean reaches rural hospital*. Institute of Industrial Engineers. Retrieved March 2, 2009, from http://www.iienet2.org/ Details.aspx?id=10808

Deming, E. (1986). *Out of the crisis*. MIT Press.

Emiliani, M. (2008). *Practical lean leadership: A strategic leadership guide for executives*. Kensington, CT: The CLBM, LLC.

Evans, J. R., & Lindsay, W. M. (2005). *An introduction to Six Sigma & Process improvement*. Cincinnati, OH: Thomson Southwestern Publishing Company.

Farris, J., Van Aken, E., Doolen, T., & Worley, J. (2009). Critical success factors for human resource outcomes in Kaizen events: An empirical study. *International Journal of Production Economics*, *117*(1), 42–65. doi:10.1016/j.ijpe.2008.08.051

Food and Agriculture Oorganization. (1999). Guidelines for quality management in soil and plant laboratories. *FAO Soil Bulletin, 74.* National Resources Management and Environment Department. Retrieved June 28, 2011, from http://www.fao.org/ docrep/w7295e/w7295e04.htm

Harris, G., & Donatelli, A. (2005). *Value-stream mapping as a lean management tool*. Huntsville, AL: University of Alabama.

Isam, E., McWilliams, D., & Tetteh, E. (2010). Enhancing lean manufacturing learning experience through hands-on simulation. *Simulation and Gaming, 41*(4). Sage Publications.

Karen Martin & Associations. (2011). *Lean terminology*. Retrieved June 28, 2011, from http://www.ksmartin.com/ resources/terminology.html#f

Laraia, A., Moody, P., & Robert, R. (1999). *The kaizen blitz: Accelerating breakthroughs in productivity and performance*. John Wiley & Sons Inc.

Liang, K. (2010). *Aspects of quality tools on total quality management*. Academic Search Complete.

Martin, K., & Osterling, M. (2007). *The kaizen event planner: Achieving rapid improvement in office, service, and technical environments*.

Mika, G. (2005). *Kaizen event implementation manual*. Society of Manufacturing Engineers.

Miller, J. (2005). *Introduction to lean and value-stream mapping*. Gemba Research.

Quick Changeover. (2010, August 23). *A lean journey.* Retrieved June 28, 2011, from http://www.aleanjourney.com/ 2010/08/quick-changeover.html

Shingo, S. (1985). *A revolution in manufacturing: The SMED system*. Productivity Press.

Smith, R., & Hawkins, B. (2004). *Lean maintenance: Reduce costs, improve quality, and increase market share*. Elsevier Science Life Cycle Engineering Series.

Soderquist, K., & Motwani, J. (1999). Quality issues in lean production implementation: A case study of a French automotive supplier. *Total Quality Management*, *10*(8), 1107–1122. doi:10.1080/0954412997091

Tapping, D., Luyster, T., & Shuker, T. (2002). *Value stream management*. New York, NY: Productivity Press.

Tpslean. (2011). *Throughput solutions*. Retrieved June 28, 2011, from http://tpslean.com/ glossary/fifodef.htm

Vorne. (2011). *Key performance indicators (KPIs)*. Retrieved June 28, 2011, from http://www.vorne.com/ learning-center/kpi.htm

Widyadana, G. A., Wee, H. M., & Chang, J.-Y. (2010). Determining the optimal number of Kanban in multi-products supply chain system. *Journal of International of Systems Science, 41*(2). Taylor & Francis.

Womack, J. P., & Jones, D. T. (1996). *Lean thinking* (2nd ed.). New York, NY: Simon & Shuster.

Zimmerman, B. (2010). What do we mean by empowerment? *Policy & Practice*, 68.

KEY TERMS AND DEFINITIONS

First In First Out: Is an inventory management strategy enacted to ensure that quality is upheld on all shipped products and components.

Global Supply Chain: Refers to the network created among different worldwide companies producing, handling, and distributing specific goods and/or products.

Kaizen Event: Is a short, focused project aimed at eliminating production waste in a particular area of an organization.

Lean Manufacturing: Is a manufacturing strategy that seeks to produce a high level of throughput with a minimum of inventory.

Metrics: Meaningful measures that target continuous process improvement actions.

Quality: Is the customer's perception of how a good or service is fit for their purpose and how it satisfies stated and implicit specifications.

Standard Operation Procedures: Is a document containing instructions on how to perform a task. It ensures that routine jobs get performed safely and in compliance with applicable regulations.

This work was previously published in Customer-Oriented Global Supply Chains: Concepts for Effective Management, edited by Ephrem Eyob and Edem G. Tetteh, pp. 224-235, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 66
Direct Building Manufacturing of Homes with Digital Fabrication

Lawrence Sass
Massachusetts Institute of Technology, USA

ABSTRACT

Architecture, engineering, and construction industries maintain a long standing desire to enhance design communication through various forms of 3D CAD modeling. In spite the introduction of Building Information Modeling (BIM), designers and builders expect varying amounts of communication loss once construction has started due to indirect construction techniques or hand based methods to manufacture buildings. This is especially true for houses and small structures, buildings that makeup the core of villages and suburbs. Unfortunately, paper documentation and reading 3D CAD models on screen continue the trend of indirect production defined in most manufacturing industries as error. The emerging application of CAD/CAM within design and construction industries provides hope for elevated communication between design and building. With CAD/CAM, it is possible to manufacture buildings of all types and sizes directly from CAD files similar to mass produced artifacts, thus reducing complexity in communication between parties. This chapter is presentation of one process of direct manufacturing from CAD and the emerging possibilities for small building production using digital fabrication. The chapter will focus on houses to illustrate the potential of direct manufacturing of buildings from CAD data.

1. INTRODUCTION

For centuries, architects and builders have pursued systematic ways to design and deliver homes at low cost in production and high quality in output. New arguments around home production are directed at machine based manufacturing of buildings opposing common handcraft construction techniques (Kiernan & Timberlake 2005). This process, typically described as prefabrication, is a century old, westernized system of home production in factories by assembly of large units with cranes on site. It has survived many decades of reinvention while also struggling for

DOI: 10.4018/978-1-4666-1945-6.ch066

broad acceptance as a worldwide industrialized system of building production (Davies 2005). In particular prefabrication has not caught on in developing countries or as a means to produce low cost housing in westernized environments.

This chapter argues prefab is complex in production maintenance, it suffers from the limitations of space and skilled labor, both impede the number of modules produced daily when demand for prefabricated homes is high. These factors also jeopardize the financial security of each manufacturer when unit sales are low. Also noted is the high cost of factory startup and operations curtailing new ventures within impoverished countries and states.

Digital fabrication or CAD/CAM is emerging as the next method of building delivery with built examples as experimental and exotic structures (Sass 2006) (Iwamoto 2008). It is a low cost, high precision manufacturing from CAD data commonly used today for furniture manufacturing. Specific to home construction one benefit of digital fabrication is that it can expand the range of production beyond the local contractor. With digital production manufacturers of metal, stone, plastic and wood based trades can fabricate components anywhere and at anytime. Manufactured components can be delivered to the project site or factory supporting assembly only production systems with low skilled labor. Digital fabrication of components assures component assembles with few discrepancies between the digital representation in CAD and the physical artifact. As a production system it allows for replication and recombination of digital models potentially supporting mass customized home manufacturing (Duarte 2005). Best is that digital fabrication can increase the efficiency of Building Information Modeling (BIM) by empowering *direct building manufacturing* from CAD based product models (Eastman 2008).

The grand challenge for *direct building manufacturing* will be discovery of new process pathways that bridge the physical divide between 3D building information models and machine data. This limitation stems from industry maintenance in hand tooling and assembly with handheld machinery as the core production method. For wood framed housing the standard material is dimensional lumber processed with standard tools such as power saws, screwdrivers and nail guns. Power tools and dimensional lumber do not take advantage of the efficiency of CAD modeling or CAD/CAM machinery. Precise machine cutting of is often difficult without special machine setups and rigging, the industry of wood framed housing maintains a need for highly skilled crafts people.

The aim of this chapter is to provide a context for digital fabrication as a mode of building production from a design by presentation of a process with two built examples. The chapter starts by presentation of industrialized manufacturing as the background for our home product manufacturing goals (Section 2). Next, past methods of controlled home manufacturing is presented illustrating the limitations and error in construction leads to higher cost and lower quality construction (Sections 3-5). Digital fabrication is presented as a method of production along with and explanation of its limitations (Sections 6-7). Materializing Design is systematic way to compute a design model for digital fabrication along with illustrations of the transformation process (Sections 8-9). The chapter ends by discussion on next steps as they related to the integration of other building systems such as plumbing, electrical, solar, etc.

2. INDUSTRIALIZED MANUFACTURING

As we stand at the threshold of new demands for energy-efficient green homes, we are also faced with the need for advanced systems that control design and home delivery. More than ever before, consumers expect new homes to perform in ways similar to mechanical products like automobiles, computers, and airplanes. Demands for increased building performance, reduction in cost, and speed in delivery all drive the desire for homes to be built in controlled environments.

The American housing industry supports three common home-delivery systems. The first—wood framing on site with dimensional lumber—is considered to be the least controlled by the designer. Here methods of production are governed by the local contractor and vary between contractors. The second form is factory-built homes as an assembly of framed boxes or prefabricated housing. Units are designed and manufactured in a factory setting as a way to control labor and materials processing. Last is a panelized system. The process starts with metal or timber framing infill with wood or metal panels. Both panelized and box fabrication require delivery by truck and on-site assembly with the assistance of cranes. All of the three methods are used by many home manufactures and designers.

The production of a prefabricated house starts with the customer's selection of predetermined designs from a catalog of glossy photos and renderings. Design styles are limited to rectangular shapes for easy transportation; cylindrical and nonplanar building shapes are uncommon. Once the order is placed, designers can then tailor each home to meet the environmental and regulatory needs of the town and site.

The efficiency of prefabricated production is its ability to standardize the mode of production. For example, the structural framing of most homes is assembled from dimensional lumber and plywood. In spite of the standards, though, every factory-made home requires some number of design changes due to variations in building site, codes, utility hookups, and the foundation. Therefore, paper drawings are generated for each new home project to guide the workers through production on the factory floor. After factory fabrication, local contractors are responsible for attaching the prefabricated boxes or panels to the foundation, as well as adding exterior siding, electrical and plumbing hookups, roofing, and finishes.

Unlike in the building industry, the design and production of goods in other major industries (e.g., automobile, computer, cell phone, aeronautical) is made up of *Original Equipment Manufacturers* (OEMs) whose responsibility is based on the assembly of products in factories (wheels, body panels, etc.), the result of which are branded products (Herbig & O'Hara 1994). The definition of OEM is not clear; each industry has its own terms and definition. In this chapter "OEM" refers to the producer of the home. Unlike home manufacturers, the assembler physically produces a minority of the assembled components; rather, they design, integrate parts, and assemble purchased products.

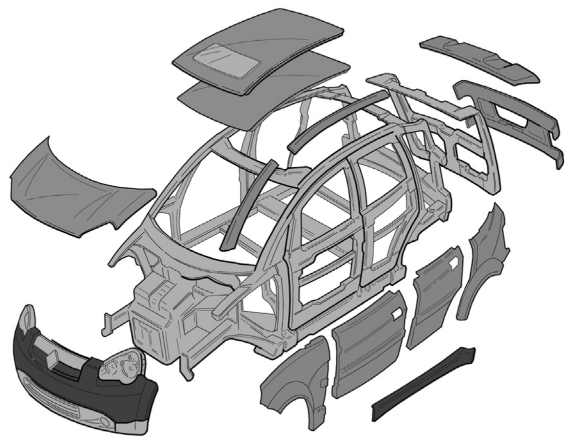
Digital fabrication makes this type of production possible by including the same or similar CAD data used in design as part of the machine manufacturing and assembly process. In the automobile industry, for example, integrated CAD systems are a critical part of manufacturing plastics and thin metal auto-body parts (Stauber & Bollrath 2007). Efficient production of plastic auto-body parts from a variety of vendors is possible because the design models from the assembler are shared by the designer and plastics fabricators (Figure 1).

Prefabricated-housing factories aspire to operate and to manufacture homes in ways similar to the OEM model. They fall short of this goal, however, due to wasteful processing and low precision. Although it is not a typical constraint in building construction, precision in manufacturing allows for the perfect fit between parts. Once high precision is available, manufacturers can have subcontractors compete to provide the factory with precisely fitting components at the lowest possible cost.

3. CONTROLLING DESIGN WITH DRAWINGS

A common theme in design control over quality, cost and time is controlled through documentation such as instructional books, CAD models,

Figure 1. Exploded diagram of a CAD model for a car produced by an automobile assembler. The chassis is built by one manufacturer, whereas panels, dashboard, and mechanical components are built by other manufacturers and then factory-assembled.



and design automation. One standing tradition is the control of style and process with rules illustrated on paper a method that dates back to the architectural treatise: Vitruvius (Vitruvius 1960) Palladio (Palladio 1965), Alberti (1986), and Serlio (1996). The seventeenth-century architect Durand pursued a similar interest through his *Précis des Leçons d'Architecture*, published in 1819—a text illustrating ways to build public buildings and statues (1802). A current form of design control can be found in architectural pattern books, which guide the production of styles and functions at the scale of a town or community (Gindroz & Robinson 2004). It is important to note that pattern books do not propose individual designs as an assembly of elements, such as columns, stairs, or window and door details, the way a treatise might. Instead, pattern books detail ways of organizing the physical elements of a community (streets, statues, street signs, etc.).

Contemporary design control as an automated process has been explored through shape-grammar methods and generative systems (Stiny 1980). A shape grammar produces designs in a certain architectural language via the application of shape transformation rules. These transformational rules allow for shape evolution from an initial shape. One example of shape-grammar production is the Queen Anne Grammar developed by Ulrich Flemming. The drawings that resulted from the grammar were used to produce designs for a pattern book as part of Pittsburgh's redevelopment initiative in the 1980s (Flemming 1987) (Flemming et. al 1986). More recently, a modern shape grammar called Malagueira Grammar was used to generate house designs in the style of Alvaro Siza (Duarte 2005). Both housing styles—Queen Anne and Malagueira—were generated by a grammar that resulted in 3D representation as drawings or CAD models. One valuable reason to consider shape grammars as a method of design generation is that the process can be automated. Unlike generating designs one at a time through keyboard entry of commands into a CAD program, an automated shape grammar program can, in minutes, generate hundreds of alternative design schemas based on a particular style. The customer can then select the best fit.

Unfortunately, all three examples (treatises, pattern books, and shape grammars) only demonstrate a way to produce design concepts and details as drawings for interpretation by builders. They do not control physical production, cost, or final product quality. Once the design is complete, all of these methods still require many laborious physical steps to build a home.

4. CONTROLLED DELIVERY WITH MACHINES

Alternatively, there are those who venture control over production by combining design and delivery. Architects and builders such as Karl Strandlund, Walter Gropius, Frank Lloyd Wright, and Konrad Wachsmann built delivery systems they hoped would make well-designed houses affordable. They collapsed the extensive relationship between the designer and contractor into companies that provided both.

Sears and Roebuck started the process through the sales of traditional and arts-and-crafts home styles from their catalog in 1895. They brought craft to the masses with a factory-based delivery system that employed artisans who had once built fine products from furniture to houses. A delivery system for housing was invented that gave access to customer-driven home design and delivery; no two houses were alike (Sears 1990) (Cooke & Friedman 2001). Variation was a major factor in sales and marketing. In time, over 400 designs, in many traditional styles, were offered to customers. Sears and Roebuck employed many artisans to manufacture components with handheld tools and machines in a well-lit factory. For efficient delivery, lumber pieces were precut and labeled before shipping by U.S. rail. The resulting assembly on-site was a balloon-framed structure clad with a variety of materials from shingles to brick. Sears and Roebuck survived as a home producer until the late 1930s, when it began offering financing as part of home sales, which unfortunately led to the company's destruction. Sears and Roebuck had positive sales combined with a negative cash flow.

By contrast, Walter Gropius, founder of the Bauhaus, an influential school of art in the twentieth century, was interested in a standardized, rational approach to design and home delivery (Gropius 1956). It was his way of countering crate-based variation in cost and quality due to variations at the hand of the artisan. His interest in moving past craft-based delivery systems to machine-based production was shared in the United States by his business partner, Konrad Wachsmann; together they started the General Panel Company in 1942 (Wachsmann 1961). Their product—The Packaged House—was a modest modern-style building expected to be factory built, delivered by truck, and assembled on-site with no postfactory production. It is clear in Wachsmann's writing that he thought machines and mass production were a means of controlling quality and cost. The machines enabled some automation in production. Wachsmann focused on the concept of element modularity as a way of controlling assemblies between machine-produced metal components. The Gropius–Wachsmann standardized building system considered ways of repeating the manufacture of one type of joinery with variations in panel production. They worked with standard metal machines of the 1940s set up in large spaces ready for building the components of many houses at one time. Unfortunately, the company was reduced to building only doors, and after five years of startup and preparation, it was unable to secure any housing contracts.

Finally, a notable contribution to the industry of home delivery was the Lustron House developed by Carl Strandlund in 1945. His houses were also factory-based products that took advantage of industrial machining and new materials processing. His company produced 2,500 homes from a handful of designed models available in four different color finishes. Exterior and interior walls were manufactured steel panels with a porcelain finish, assembled on-site with nuts and bolts (Knerr 2004). Each house was assembled from over 30,000 small components, which required many manhours in labor. In fact, the manhours needed to screw bolts and gaskets drove the house price out of the affordable range. The company lasted only nine years. It was not the concept of precise machinery in a comfortable controlled environment (i.e., the factory) that made the Lustron a short success. Its demise was due to a laborious assembly system and inflexible panel making that limited variations between designs.

Today's best prefabricated systems resemble the processes employed by Strundland and Wachsmann/Gropius—hand based, hand tooled with inflexible machinery—with the exception of a few factory-based methods in Japan (Noguchi 2003) and Europe and the use of digital design. Prefabricated housing companies continue the legacy of low-precision, high-cost production at the hand of the artisan. A notable step forward in production-based research can be found in a system that integrates a structural system into the design computing system (Benros & Duarte 2009).

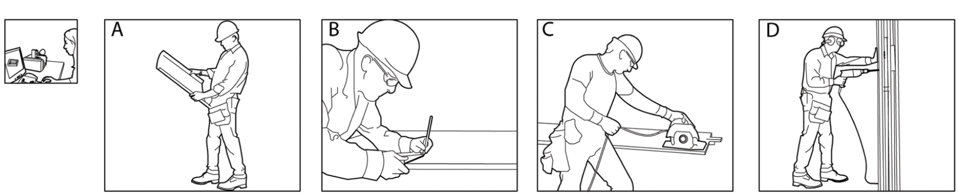
5. ERROR AND THE ARTISAN

As mentioned earlier, this is not the case in the manufacturing of common industrialized products—such as cell phones, computers, and automobiles—that use the OEM model. In any product design process, prediction and elimination of error is a key activity in the design process (Goh et al. 2005). In contrast to product design, error in home construction originates from overlapping operations between construction workers and the process of visual interpretation of the drawings. This conflict is due to measuring and cutting components with manual tools, then re-measuring previously assembled building elements and new cutting. In construction, design and shop drawings merely illustrate "design intent." Workers can and do deviate from notations on the drawing due to unforeseen obstacles in the field. For example, the process steps in Figure 2 show a designer who produces data output as drawings. These are interpreted by construction workers (a). The information from the drawing is manually transferred to a material substrate (b), hand cut (c), and assembled with handheld tools (d). This low-precision process requires time for measuring and cutting each new element. Steps c, d and e in Figure 2 are then repeated for each new element (as when fabricating a stud wall), so time is needed in order to measure previously assembled studs and to deal with any inaccurate cutting of new studs. As the complete artifact is assembled, the elements may well have little or no relationship with the designer's drawings and models. In craft-based construction, every component is hand measured, hand cut, and assembled with little dimensional relationship with the original drawing.

6. CREATIVE USE OF DIGITAL FABRICATION

For higher precision in cutting and assembly, CAD drawings can be sent to a digital fabrication device for precise measuring and cutting. Digital fabrication is a flexible three-part process from digital modeling to a fabrication device and hand or machine assembly that brings together a relationship between materials, machining and CAD modeling. In contrast to the artisanal work discussed in the previous section, digital fabri-

Figure 2. Traditional craft-based home production starting with CAD drawings



cation is computer-based measuring, computer-controlled cutting, and hand assembly of precise components. An extensive explanation of digital fabrication machinery, machining methods with examples is found in Schodek (2007). There are two flavors of digital production for the field of architecture.

The first being project based digital design and digital manufacturing. Digital fabrication in the field of architecture is defined by architects who create one-of-a-kind buildings as sculptural artifacts. Design commissions completed by architects such as Frank Gehry who fabricates buildings of metal and wood, theorist Greg Lynn and Bernhard Cache postulates ideas about fabrication through prototyping. Lisa Iwamoto, Nader Tehrani and Mark Goulthorpe are proof of the creative potential behind these new technologies by fabrication free form exhibitions of fabrication. These architects represent their designs with advanced curved surfaces as well as solid and parametric modeling software, and then physically produce complex designs by CAD/CAM manufacturing. They employ a combination of sophisticated technologies such as curved surface CAD modeling tools to simplify areas of design complexity. For most design and construction professional direct manufacturing is reserved for a few areas of construction such as building exterior only, interior surfaces, and the structural members only. A summary of one-of-a-kind projects and methods of production including both traditional and CAD/CAM fabrication can be found in Bernstein (2010). One built example of a creative design can be found in the Amsterdam Pavilion in lower Manhattan built of digitally fabricated plywood components. This project was modeled in CAD traditionally with key board entry of modeling commands. Components and assembly are with metal fasteners and a hand finished surfacing (www.unstudio.com). An example to follow is a project built by the Institute for Advanced Architecture Catalonia of a solar paneled house. This project is also a digitally fabricated plywood structure with assemblies sustained by metal fasteners (www.iaac.net/projects/fab-lab-solar-house-3).

The second type of digital fabricated architecture is as a production system where multiple iterations of a design are generated and manufactured. The first example Contour Crafting is one example of rapid digital fabrication of houses by using layered concrete dispensed from a computer-controlled machine. The greatest potential of Contour Crafting relates to housing in developing countries in need of original and replacement concrete structures (Koshnevis 2004). A second example of a production system is defined as materialization. It is a systematic method of subdividing geometry into constructible components ready for digital fabrication. A pilot project built summer of 2005 demonstrated the potential of digital fabrication as a system of building production. The project was a small cabin built completely of interlocking plywood components manufactured from CAD/CAM machines. The structure measured of 8' x 10' x 15', was elevated 24" off the ground. A detailed description of the process and computation can be found in Sass (2005) (2007). A second example of a materialized design was also constructed using from interlocking plywood components three year later. This version was constructed as part of a museum exhibition was assembled with few tools and assemblies that were sustained by friction from part interlocking (Bergdoll 2008). From these two examples of digitally fabricated structures it is possible to mass produce many of these buildings from the same data. It also means manufacturing is possible anywhere using similar machines controlled by computers. A series of construction rules were developed from the two examples.

Digitally Fabricated Structures:

1. Computer generated designs are manufactured from computer controlled machines
2. A variety of complex components can be manufactured by the machine

3. The structure is built of layers of materials that compose a lattice & surface
4. Each element includes interlocking geometry for direct association with other components
5. Elements are fabricated as 2D parts
6. Parts are assembled with few tools possible to sustain by friction

7. PROBLEMS WITH DIGITAL FABRICATION

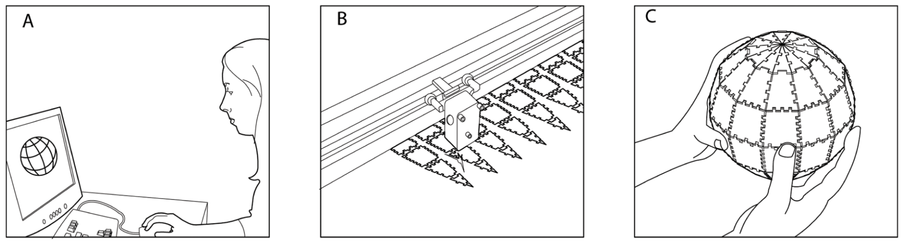
A core problem in digital design and fabrication is the many design steps associated with modeling physical objects in CAD as described by Sass and Oxman (2006). The paper outlines three limitations in design production: complexity in component design, generating assembly descriptions in CAD and scaling models from physical prototypes to full scale construction. Design automation can address these problems with computer programs built for rapid model generation. Researchers in electronic design automation use CAD tools and libraries to design and generate information for printed circuit boards (Jansen 2006). For digital fabrication automated CAD systems will aid in conversion of 3D design models to 2D data for digital fabrication CAD/CAM. An integrated process of design generation will elevate the impact that computational analysis and optimization tools have on designs and fabrication.

8. MATERIALIZING A DESIGN

A systematic process of object production is discussed here as a pathway towards building physically based automated design tools. This process, defined as *materializing* is intended for production of many 3D artifacts, the process assumes each artifact to differ in shape or composition of shapes with the same physical language of production. This system of production is alternative to methods used by design architects to build a kind of sculptural artifacts. *Materializing* is a computational production system that allows designers to rapidly realize visual and physical goals from three-dimensional CAD models. In summary, *materialization* is a three-stage process of shape transformation from (*a*) an initial shape model to (*b*) CAD/CAM manufacturing and (*c*) assembly by hand or with robots (Figure 3). The contribution of this system defines the mathematical relationship between the designer's CAD model and the physical building production—the critical step in the relationship between computer-aided design to computer-aided manufacturing. As for buildings a close relationship between flexible CAD modeling and flexible digital fabrication will lower process and material waste by removing process steps commonly found in craft based construction. The work in this chapter is not automated with CAD tools as mentioned above, however the format and steps provide a format for building product automation.

Alternative to the three part physical shape production illustrated above, building production requires many steps in translation. With computation quality and time is controlled by converting steps that would otherwise be handcrafted steps to computable steps. Figure 4 is an illustrated process summary that shows three of many computational steps in converting a design model generated in CAD to sets of machine tool paths. Each pathway assists full scale assembly of components with computable features between components. The first step captures geometric characteristic of the starting Shape for structural analysis and data storage (Figure 4). If the shape is structurally sound the next steps are focused on structural contours along the three axes (x, y & z) or a new shape is generated and edited for structural compliance. The new shape is composed of structural contours as interlocking geometries of varying depths depending on loading and location. The second step is the addition of attachment features to each contour, these features join the surface geometry with the contour geometry. Here

Figure 3. Three step materializing process from a shape model (a) in CAD to laser cutting (b) and hand assembly (c)



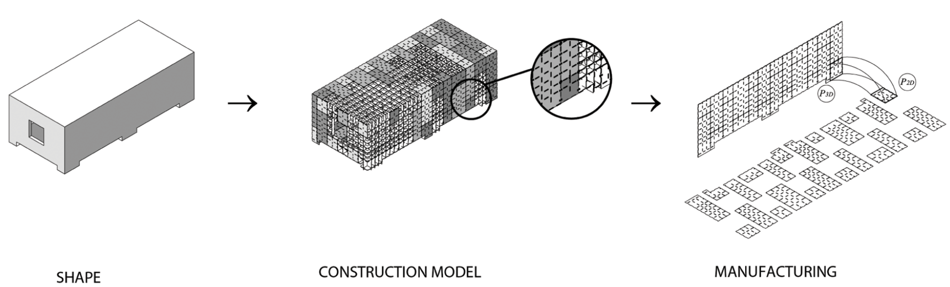
the assembly alignment and final attachment between each element is an integral part of the elements geometry. Integral attachments of this type were popularized by research in plastics as a way to measure and manufacture plastic fasteners. The fourth step in the process is subdivision of exterior surface and contours into smaller parts related to the machine and material constraints. The resulting geometric model in a finished collection of elements defines a construction model (*Construction model)*). The last function is translation or development of three-dimensional elements to two dimensions *(*Manufacturing)a function is common in the field of descriptive geometry. Developed elements and their attachment features are packed and sorted in two-dimensions before final CAD/CAM fabrication. Construction modeling of the type in Figure 4 is complex, labor intensive CAD modeling through keyboard entry of commands.

9. CONCLUSION

This chapter starts by descriptions of houses as manufactured products designed digitally yet manufactured by hand. The claim is that a physical disconnect between tools in design and physical construction lead to errors in production unfortunately considered the rule and precision based construction is the exception. This imprecision in production results in many process steps in manufacturing from hand measuring and re-measuring by hand of previously built components.

Digital fabrication for home construction introduces precision in manufacturing, assembly and component integration. Precision manufacturing throughout the design and construction process elevates the impact and purpose of computation analysis tools and design optimization. The resulting components can be assembled on site with low skilled labor due to semantic nature of each piece.

Figure 4. Broad steps in materializing a model of a building from a shape model to a series of interlocking components



Beyond this chapter, a search for holistic approaches to building design, computation and production requires further development in methods to generate component descriptions. In this text computational systems decompose an initial shape into sets of components for digital fabrication in ways similar to carpenters on site. This system also illustrates how the shape of a building can be analyzed and optimized for assembly and fabrication controlling cost and the quality of design delivery. Illustrated are abstract steps in construction modeling for the internal structure, external surfacing and final production data. It is considered first steps leading to design automation for digital fabrication only for the structure of a building.

Future work will seek integration of other building systems beyond structure into interrelated systems such as rain-screens, insulation, air-handling, electrical, plumbing, etc. It is believed that most if not all of these systems can be fabricated through CAD/CAM design as well. It also means that each of these systems requires re-invention and development beyond the three dimensional methods of common construction towards 2D productions that assemble into 3D buildings. New research in digitally based construction systems will lead to similar breakthroughs in construction where all systems can be designed rapidly in CAD and manufactured with machines.

ACKNOWLEDGMENT

The author thanks Dennis Michaud and Daniel Smithwick for their dedicated service in building the exhibit structure, the Department of Architecture at MIT, and the Rockefeller Foundation, which was the primary sponsor of the MoMA exhibit. Additional thanks go to the American Institute of Architects and Boise Cascade, as well as to Shopbot Tools for fabrication.

REFERENCES

Alberti, L. (1986). *The ten books of architecture*. New York, NY: Dover Publications.

Benros, D., & Duarte, J. (2009). An integrated system for providing mass customized housing. *Automation in Construction*, *18*, 310–320. doi:10.1016/j.autcon.2008.09.006

Bergdoll, B., & Christensen, P. (2008). *Home delivery: Fabricating the modern dwelling*. New York, NY: Museum of Modern Art.

Bernstein, P., & Deamer, P. (2010). *Building (in) the future: Recasting labor in architecture*. Princeton Architectural Press.

Cooke, A., & Friedman, A. (2001). Ahead of their time: The Sears Catalogue prefabricated houses. *Journal of Design History*, *14*(1), 53–70. doi:10.1093/jdh/14.1.53

Davies, C. (2005). *The prefabricated home*. London, UK: Reaktion Books.

Duarte, J. (2005). Towards the mass customization of housing: The grammar of Siza's houses at Malagueira. *Environment and Planning B*, *32*(3), 347–380. doi:10.1068/b31124

Durand, J. (1802–1805). *Leçons d'Architecture Partie Graphique des Cours d'Architecture*. Paris, France: Chez l'Auteur.

Eastman, C., Teicholz, P., Sachs, R., & Liston, K. (2008). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers, and contractors*. New York, NY: John Wiley & Sons.

Flemming, U. (1987). More than the sum of parts: The grammar of Queen Anne houses. *Environment and Planning B*, *14*(3), 323–350. doi:10.1068/b140323

Flemming, U., Coyne, S., Pithavadian, R., & Gindroz, R. (1986). *A pattern book for Shadyside: Technical report*. Department of Architecture, Carnegie-Mellon University, Pittsburgh, PA.

Gindroz, R., & Robinson, R. (2004). *Urban design associates: The architectural pattern book*. New York, NY: W.W. Norton & Company.

Goh, Y., McMahon, C., & Booker, J. (2005). Development and characterization of error functions in design. *Research in Engineering Design*, *18*, 129–148. doi:10.1007/s00163-007-0034-x

Gropius, W. (1956). *The new architecture and the Bauhaus*. Boston, MA: Charles T. Branford Company.

Herbig, P., & O'Hara, B. (1994). The future of original equipment manufacturing: A matter of partnership. *Journal of Business and Industrial Marketing*, *9*(3), 38–43. doi:10.1108/08858629410066854

Iwamoto, L. (2008). *Digitally fabrications, architectural and material techniques*. New York, NY: Princeton Architectural Press.

Jansen, D. (2006). *The electronic design automation handbook*. Kluwer Academic Publishers.

Kiernan, S., & Timberlake, J. (2005). *Refabricating architecture: How manufacturing methodologies are poised to transform building construction*. New York, NY: McGraw-Hill.

Knerr, D. (2004). *Suburban steel: The magnificent failure of the Lustron Corporation, 1945–1951*. Columbus, OH: The Ohio State University Press.

Koshnevis, B. (2004). Automated construction by contour crafting-related robotics and information technologies. *Construction and Automation*, *13*(1), 1–19.

Noguchi, M. (2003). The effect of the quality-oriented production approach on the delivery of prefabricated homes in Japan. *Journal of Housing and the Built Environment*, *18*(4), 353–364. doi:10.1023/B:JOHO.0000005759.07212.00

Palladio, A. (1965). *The four books of architecture*. New York, NY: Dover Publications.

Sass, L. (2005). A wood frame grammar: A generative system for digital fabrication. *International Journal of Architectural Computing*, *1*(4), 51–67.

Sass, L. (2006). Synthesis of design production with integrated digital fabrication. *Automation in Construction*, *16*(3), 298–310. doi:10.1016/j.autcon.2006.06.002

Sass, L. (2007). Synthesis of design production with integrated digital fabrication. *Automation in Construction*, *16*(3), 298–310. doi:10.1016/j.autcon.2006.06.002

Sass, L., & Oxman, R. (2006). Materializing design, the implications of rapid prototyping in digital design. *Design Studies*, *26*, 325–355. doi:10.1016/j.destud.2005.11.009

Schodek, D., & Bechthold, M. Griggs, K., Kao, K., & Steinberg, M. (2007). *Digital design and manufacturing: CAD/CAM applications in architecture and design.* New York, NY: John Wiley & Sons.

Sears & Roebuck. (1990). *Home builders catalogue, The complete illustrated 1910 edition*. New York, NY: Dover Publications.

Serlio, S. (1996). *Sebastiano Serlio on architecture: Books I–V of Tutte l'Opere d'Architettura et Prospetiva*. New Haven, CT: Yale University Press.

Stauber, R., & Bollrath, L. (2007). *Plastics in automotive engineering: Exterior applications*. Munich, Germany: Hanser.

Stiny, G. (1980). Introduction to shape grammars. *Environment and Planning B*, *7*(3), 343–35. doi:10.1068/b070343

Vitruvius, M. (1960). *The ten books on architecture* (Morgan, M. H., Trans.). New York, NY: Courier Dover Publications.

Wachsmann, K. (1961). *The turning point of building*. Wurzburg, Germany: Reinhold Publishing Corporation.

KEY TERMS AND DEFINITIONS

BIM: Generation and management of building data, typically a 3D model, throughout its life cycle representing spaces, products and some means and methods of production. The model can be updated for changes made between design, finished construction and renovation.

Construction Modeling: 3D component based description of a design in CAD. This description challenges component at the scale of assemblies. Each digitally fabricated component is represented in 3D.

Design Modeling: 3D description of a design in CAD. This description is a formal representation used for visual evaluation of a design.

Development: Development follows the term developable, as in geometry that can be flattened to 2D from a 3D shape.

Digital Fabrication: Direct manufacturing of physical artifacts from CAD files.

Layered Manufacturing: Similar to additive manufacturing where artifacts are produced from layers of material. Each wafer thin layer is manufactured from a CAD model and automatically assembled to the previous layer.

Materialization: Translation of a 3D design model into 2D elements for CAD/CAM manufacturing.

OEM: Original equipment manufacturer is the company that originally manufactured a specific product.

Rapid Prototyping: Automated production of a physical 3D object from a CAD description.

This work was previously published in Computational Design Methods and Technologies: Applications in CAD, CAM and CAE Education, edited by Ning Gu and Xiangyu Wang, pp. 71-82, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 67
eRiskGame:
A Persistent Browser-Based Game for Supporting Project-Based Learning in the Risk Management Context

Túlio Acácio Bandeira Galvão
Rural Federal University of the Semi-Arid – UFERSA, Brazil

Francisco Milton Mendes Neto
Rural Federal University of the Semi-Arid – UFERSA, Brazil

Mara Franklin Bonates
Rural Federal University of the Semi-Arid – UFERSA, Brazil

ABSTRACT

Motivated by the increasing demand for software engineering professionals, in particular project managers, by the dissemination of the use of games as an attractive instrument in the learning process and by the universalization of the Web platform as a catalyst of human relations nowadays, this chapter proposes the use of a Persistent Browser-Based Game as a support in the qualifying process for new professionals of Project Management. Following the pedagogical theory of the Project-Based Learning - PBL, the game gives the player the opportunity to experience real situations of Project Management by proposing challenges commonly faced in most enterprises. These ever present challenges include unpredictability in the software manufacturing organizations, by means of the use of intelligent agents to assign challenges and common barriers to Software Projects.

INTRODUCTION

The importance and the real needs of the adoption of methods and principles of the Project Management (PMI, 2004) in organizations are currently widely discussed and accepted. In these organizations, the main role of the Project Manager is the task of conducting a project to its successful conclusion. However, that is not what usually happens in software projects (Brewer, 2005).

DOI: 10.4018/978-1-4666-1945-6.ch067

In an attempt to explain the high number of projects that fail due to reasons related to bad management, some studies have discussed a possible relationship between the lack of certain abilities by managers and the traditional teaching methods. This is one of the consequences of acquiring knowledge without an experience in some real project or complementary educational approaches, such as games and simulations.

In order to provide a solution for this deficiency, in this chapter we propose a tool to provide a new way of learning that is not only attractive, but efficient and collaborative as well. This tool meets the needs of users with distinct routines and schedules. This chapter proposes the eRiskGame tool, which is a Persistent Browser-Based Game for educational purpose. The game is about the tasks that a Project Manager must perform in an organization. Its focus will be on Risk Management, more specifically in the Planning, Control and Monitoring (budget, time schedule and software quality).

This serious game uses PBL to bring the player a way to acquire knowledge on project management, particularly in the risk control involved in this process. To that end intelligent software agents were employed in monitoring and controlling of the environment, which is in constant change and affects the professionals, the organization and its customers.

BACKGROUND

This session introduces the background and the important terms involved in the development of this tool.

Technological Aspects and Motivators

We live in a highly dynamic, diverse and demanding society, where new technologies are constantly seeking to supply the needs and take into account the particularities of that heterogeneous public, which have different time availability, locomotion and spaces. These new technologies have the fundamental role of making these differences transparent, allowing persons with distinct dispositions and skills to interact jeopardizing their performance.

In this context, we have the Web as a powerful resource in the process of teaching and learning, through which we can learn and teach in many ways, in different places and at different speeds. This technology allows the expansion and integration of knowledge in a way that could be fast, dynamic and accessible to all providing the construction/reconstruction and socialization of knowledge for a better individual, social and collective context of all involved.

There has been a growing interest in the use of computer games as a didactic-pedagogical tool, for training, qualification or improvement of skills in several knowledge areas. Not only the academy, but also the industry has shown great interest in this market portion, which already moves $20 million per year (Susi et al., 2007), and attracts more and more attention of those seeking a way not only innovative, but effective to transmit knowledge without the common lack of motivation in the traditional distance education methods.

Software Engineering and Teaching Model

In the context of software engineering, numerous initiatives have arisen in order to use the ludic qualities and motivation that the games have in the education or training of new professionals, in an attempt to reduce the limitations imposed by time and/or format of the traditional courses in this area.

In the various educational levels, one of the factors that most concern the teachers is the motivational level of their students. In Sandford et al. (2006) it has been observed that 53% of the teachers see the motivation for their pupils as primary reason for using educational games. However, the authors argue that the use of games in educational environments does not ensure motivation alone. We must look to the characteristics of the games that better encourage the students' commitment.

In addition to motivation, computer games may also be able to introduce integrated techniques, involving a wide range of interaction ways that improve both the learning and the sociability. The Project-Based Learning is a theory that has gained a special interest in the serious games area, because the students not only improve their analytical capacity, but they also assimilate more knowledge. Games also offer experience of collaborative work and improve the ability of players' communication.

PROJECT-BASED LEARNING

Project-Based Learning (PBL) was proposed by Ayas and Zeniuk (2003) as a mean to conduct the education where the personal and collective skills are devised through the development of learning capabilities which allow short term reasoning. This learning theory allows individuals and teams the creation and the sharing of knowledge. This theory presupposes that a project centered on tasks where the pressures of short term prevail, the participants must be in balance between action and reflection, in order to build the competence of the learning capacity (Hsu and Liu, 2005).

The PBL can be understood as an education technique, where situations of an actual context are modeled on a fictitious project in which the students must commit-to-finish it, and thus build knowledge regarding that experience.

A project is defined as a temporary effort, driven by a group of persons, to create a product, service or single result. By temporary we mean that all the projects have beginning and end defined. The end is reached when the goals are achieved, otherwise the project has failed. A project is characterized by having an objective and clearly defined requirements (scope), by complying with the approved budget (cost) and must be finalized within specified term (time) (Schwalbe, 2002).

In the PBL, the learning comes through a process in which the students build a product, working as a team. The product may be something tangible (such as a model / prototype, a system or a robot), a computer product (such as a software, a presentation or a multimedia product), or a written product (such as a report, an assessment or a summary of experimental results). The product must either answer a question, solve a problem, or attend the needs and requirements laid down by the course instructor or identified by the students (Frank, 2008).

PBL is a hands-on approach that is focused on issues that lead students to meet the fundamental concepts and principles of a subject. It helps their abilities of creative thinking, showing that there are many ways to solve a problem. In addition, the education based on projects differs from the traditional education by its emphasis on the artifacts construction by the students themselves to represent what is being learned.

The central idea of the PBL is that the problems of the real world can capture the interest of students and cause serious reflections on how students may acquire and apply new knowledge in a context of problem resolutions (Yadav and Xiahou, 2010; Frank, 2008). The teacher plays the role of the facilitator, working with the students for raising relevant questions. The teacher also helps in structuring significant tasks, improving both the knowledge development and the social skills. In addition, the teacher evaluates carefully what the students have learned from the experience. The supporters say that the PBL helps students to acquire thinking skills besides of the collaboration required in the workplace.

PBL Applied to Software Engineering

The traditional approach in the Software Engineering teaching is based on a reading model.

However, this model brings a big problem for students, mainly due to the little involvement with the theme. Students play a passive role in the educational process, differently of the role of Software Engineer which must be alert to what occurs in the projects which he or she coordinates and make decisions that will be the foundation for their successful conclusion.

Integrating the PBL in the learning of Software Engineering makes it possible to provide the student with a practical experience that cannot be obtained in activities performed in a laboratory (Yadav and Xiahou, 2010). In the traditional approach, the problems are normally adapted and simplified in such a way that they do not appear to be relevant or are linked to solutions already pre-manufactured, which prevent reasoning and gathering of the students' ideas to deal with problems of this nature. In addition, some issues, like process models, appear to be so theoretical that do not show the students how this will be used in practice.

Based on these facts, we see that the PBL, when applied to serious game for the teaching of Software Engineering is as feasible as useful to the students in relation to the practical that the discipline. This combination may also improve the learner's analytical capacity, which concerns the ability to analyze data, not always inter related, and from this analysis produce valuable information or knowledge.

GAME-BASED TRAINING

In the evolutionary rhythm, which the humanity has walked, many paradigms were broken and much has been gained with this. In the education context, what we have seen are new methodologies and educational approaches being created and recreated. The aim is to achieve better pedagogical results with an ever more dynamic and diversified public. One of these approaches is the use of electronic games in the educational context (SGI, 2010), which has been broadly discussed and implemented in formal and non-formal education.

Formal Education represents a set of educational activities with clear and well defined goals which conform to centralized educational directives and are implemented within the limits of schools and universities. The non-formal education is more diffuse, less hierarchical, unbureaucratical and is usually performed outside the limits of the formal educational system.

Some authors consider also the existence of informal education which, although not submitting the rigor and the implementation of traditional educational methods, promotes some means of acquisition of skills and knowledge (Sørensen and Meyer, 2007).

In the several educational models, some factors, such as lack of motivation or involvement with the subject taught often become very difficult obstacles to be transposed and may impair the student's performance. In this sense, games may be integrated to the training models by adding fantasy, visual effects and interaction way more attractive to the students.

Some of the main elements found in games include interaction, flexibility, competition, visual feedback, dramatic effects, usability and fidelity degree to reality (Prensky, 2001). Interactivity is a fundamental characteristic, because the player can easily get bored without an active participation. The flexibility provides the player with a control level and investigation of the available content, in accordance with the rules. The competition represents the challenge, the conflict, the solution of problems that require exploitation and strategic abilities that motivate the player in the search of the goals defined. The game provides feedback to the player actions, evaluating and presenting the player's progress in seeking his or her goals. This feedback is done by means of dramatic effects on the elements that compose the game story. These effects (sound, images, animation, music, humor and stories) can bring entertainment to the learning content, adding a context, characters and dramatization. That allows the student to experience a role and act within the game.

In order to use computer games and games in general for educational purposes, several aspects of the learning process must be supported: i) students must be stimulated to combine the knowledge of different areas to choose a solution or to make a decision in a certain moment; ii) students must be able to test how the game result changes based on their decisions and actions; iii) students must be encouraged to keep contact with other team members to discuss and negotiate the subsequent steps, improving, thus, among other things, their social skills (Pivec et al., 2003).

Software Engineering students in particular can benefit from the use of games in training once the project management depends strongly on past knowledge and experience. Using games students can analyze situations already experienced in other projects and evaluate different paths on the project that he or she could have taken, due to specific decisions that have been made on specific points, thereby improving their abilities on management and decision making.

BROWSER GAMES

Growth of the Web and People Interest on Browser Games

According to research carried out by the International Telecommunications Union (ITU), from 2003 to 2009, the number of people who have Internet access doubled and represents 25% of the world population, and in developed countries this number rises to 64%. Following this perspective of growth, from 2009 to 2010 the number of users that have Internet access at home have passed from 1.4 to 1.6 billion.

In view of the growth of "connected" public, it is increasingly interesting to invest in the Web platform. Along with that growth, it is possible to notice social changes experienced in the past 25 years that have arisen what is known today as Digital Natives. They are people with easy and permanent contact with technology, representing today 50% of the active population and may reach 80% in 10 years.

Aiming to reach that market portion in constant growth, the billionaire electronic games industry has started to invest in multimedia titles totally in Web, focusing on the potential that the platform offers, such as: i) persistence, which is the state maintenance of the game in the server aiming posterior access; ii) platform independence, an increasingly necessary requirement, given the diversification of operating systems and devices for Web access; In addition to the iii) collaboration experienced in Web 2.0 and the explosion of social networks.

The browser games may include all the genres of games, which could be single-player[1] or multiplayer[2], and the latter gained more emphasis due to its additional focus on the social interaction, often on a massive scale. Social networks make use of these games to bring more people, for example the Zynga's Farmville (Zynga Inc., 2011), popular farm browser game running on Facebook.

Persistent Browser-Based Game

In this context, we can define a Persistent Browser-Based Game (PBBG) as an electronic game that can be played and accessed by a Web browser and presents a shared persistent virtual environment, where the events continue to occur even in the absence of user. The user may recover his session later and continue in the game (Project, 2009).

In a heterogeneous audience, with specific requirements of time and resources, the characteristics offered by PBBGs may be the key-point in the success of a learning tool. Since it provides people with separate provisions to interact, compete or exchange experience using the same tool. People who need to move could benefit from the portability of the PBBGs, once the information about the users' profiles are stored in the server, and no special software is necessary, so that they may be accessed from any terminal through a browser and continue from the point where he or she stopped.

PBBGs can also run on mobile devices, most of these with limited capacity. This is possible because most part of browser games do not count on complex graphs or sounds, which normally are compensated with one ludic aspect, thus maintaining their attractiveness.

Due to the frenetic emergence of new multimedia titles, in many styles, qualities and themes, it has been avoided in this work cite examples. However, we can indicate Browser Games (2010) as a wide and detailed list of the main browser games currently available.

WEB COMMUNICATION TOOLS APPLIED TO EDUCATION

Advances in communication technology are changing the way in which people around the world teach and learn. The communication technology offers new alternatives for creation, storage, access, distribution and sharing of didactic material (Chou and Sun, 1996). Since the Internet has become part of the life of most part of students, many Web communications tools are being exploited as an element that aids in the learning process.

Currently, a type of tool widely used in virtual activities and discussion groups is the forum. In an asynchronous interactivity, participants may exchange views and discuss themes proposed, generating an exchange of experiences and questions which motivate users to deepen in problem context and search for more knowledge, since the diverse opinions may cover subjects besides the discussion focus.

The electronic mail is a more direct communication tool. It is normally used for receiving guidelines directly from the teachers and may be applied to provide a private dialog between the parts. Chats offer a synchronous communication between them and allow the creation of discussion rooms separated by issues or user groups. These and other tools aim to fulfill the human need to interact with other persons and not always are included in distance education.

The social interaction promoted by communication tools via the Internet brings other benefits to traditional education methods. The students feel more comfortable to expose their doubts and may receive feedback not only from the teacher, but also from other students, allowing a comparison between the various points of view. This cooperation supports and encourages the sharing of knowledge and individual experiences, bringing to the group a wide range of experience.

The wide use of these tools has given a greater flexibility and freedom to teachers and students regarding the manner in which they organize and obtain progress with their studies and has created a environment in which learning occurs how and when it is convenient (Xakaza-Kumalo, 2010). In this context, it is noticed that learning environments, mainly educational games, need communication tools to encourage knowledge sharing and promote better interaction and collaboration between students and teachers.

INTELLIGENT AGENTS

To improve the effectiveness, or even the autonomy of computational tools, some techniques of Artificial Intelligence (AI) have been employed in various areas. Due to some of their abilities, such as: behavior guided by goals, reactivity, reasoning, adaptability, learning, communication and cooperation, Intelligent Agents have gained space and become very popular in computer games.

According to Russell and Norvig (1995), an intelligent agent is any entity which may receive information from the environment where it lives by means of sensors, and act in that environment by means of actuators, in a rationally manner. In other words, they act in a correct way, tending to maximize an expected outcome.

There are several types of agents, but each one is elaborated in accordance with the environment in which they shall be inserted and the functionalities which they should provide. They may be software or hardware and are normally classified in 4 (four) basic types (Russell and Norvig, 1995):

- The simple reactive agent selects their actions based on the current perception, ignoring the perception history;
- The model-based reactive agent maintains the internal state to control the aspects of the world which are not evident in the current perception;
- The goal-based agent need a description of the current state besides some kind of information about the goals that describe desirable situations. Thus, this agent could achieve the goals;
- Some environment states can satisfy the agent, however, a state may be chosen over another. Thus, the agent based on the utility uses a performance measure which allows a comparison among different states of the world, in order to select which one will be more useful to them.

An autonomous agent set which agents cooperate among themselves aiming to solve a problem that is beyond the capacity of a single agent is considered a Multiagent System (MAS) (Pontes, 2010).

Intelligent Agents in Computer Games

As the games get more realistic, in terms of physics and graphics, the characters and the environment have evolved and became more intelligent. Normally agents are used in computer games to provide to the systems a behavior not very obvious and different in each situation, making the player change his or her strategy and improve his or her performance in order to overcome the difficulties presented.

AI is present in all game styles, more often in RPGs (Role Playing Games) where enemies can be monsters or characters similar to the player, controlled by agents. The enemies should be autonomous and need to interact with dynamic and complex environments, which require reactive behavior, planning and common sense. They need to navigate by the virtual world, requiring (a) "pathfinding"[3], (b) space and (c) temporal reasoning. Advanced agents may have mechanisms to adapt to the strategies of their opponents, and can also learn.

Intelligent Agents may be used in games with many purposes, not only in the representation of opponents or partners, but also in the representation of the environment itself, since the virtual environments try to represent in the real environments the most accurately possible way. These environments are subject to the most diverse circumstances, like climate changing or disasters, among other things that can influence the game progress.

Intelligent Agents Applied to Educational Tools

The use of resources from AI aiming to support the development of educational systems, in addition to enable a great capacity for adaptation to context and personalization of the environment in accordance with the student characteristics, it provides a high interactivity degree between learning ambient and the users. Among the several AI techniques that may be applied in the education field, the intelligent agents are one of them.

In the distance education, students and the facilitator are separated geographically and, sometimes, they do not know each other. That way, the agent may be employed in the detection of compatibilities in students' profiles and help the teacher in the formation of groups (Pontes and Mendes Neto, 2010). That will, among other things, facilitate the dialog between students and learning, since they already have similar skills, knowledge or deficiencies.

Some educational tools have by purpose to provide a personalized teaching, since many students have difficulty assimilating the content due to the way in which this is transmitted. Some of these systems have used the agents to monitor tests and dialogs among students to analyze whether there is any learning. Otherwise, they select, in the knowledge base, another media to expose this content, adapting it to the student.

As seen, intelligent agents have been widely used in educational tools due to their flexibility and mainly their autonomy and capacity to sense and interact with the ambient. Thus, in this work, intelligent agents were used to monitor the players' projects and create challenges; in addition, it makes the environment closer to a real environment, by adding unpredictability and uniqueness to each project. The application will be better exposed in the session dedicated to the game description.

RELATED WORKS

As related works, we can cite some initiatives carried out by researchers in the software engineering area, such as TIM: The Incredible Manager (Dantas et al., 2004), a simulation game with focus on Planning and Control (budget, time schedule and quality) on the Java platform.

We also have the SimSE (Navarro and Hoek, 2004), whose focus is the software project management, including the phases of analysis, design, construction and tests; it is also developed on the Java platform. Another example is the SESAM (Drappa and Ludewig, 2000), which approaches the development software process focused on quality assurance.

E-RISKGAME

This session brings to the reader the serious game developed according to the themes presented here, in addition to the module developed for the teachers, where these evaluate the attitudes taken by players in accordance with the game progression.

Scenario and Storyline Game

The eRiskGame was created to simulate an experience in managing software projects, where the player can get ready to control the expenses, to comply with the targets and deadlines laid down, in addition to accompany the team work productivity. In view of that, as a real company, the work will be subject to changes in the administration of the organization, requiring changes by customers and other risks which may prevent the success of their projects.

The game's storyline consists of a Software House that goes through changes and is seeking new project managers. At the start, the player is involved in this context, which it is presented in the tutorial form, as shown in Figure 1.

The ludic aspect tries to captivate the player, to unwittingly submit concepts and leaving him or her free to reach them when the player considers it necessary to transpose any challenge imposed by the game. Besides the tutorial, an additional material about Risk Management, with integrated questionnaires is made available to the trainee. The questionnaire results may increase the final player's score.

Once inserted in the scenario, the player can start new projects, which have budgets, deadlines and targets, varying in acceptable intervals. This provides a certain dynamics to the game, avoiding that the players get unmotivated with predictable results. Each project is divided into phases which in turn are divided into weeks and days.

Players can monitor the result of their actions by means of numbers and graphics (Figure 2),

Figure 1. Game tutorial

which gives them better prospects of success in their decision making.

In the eRiskGame, the planning of the teams and, consequently, budget is highly important. For this, the user may contract (Figure 3) the most varied professionals and fire them at the time the user may deem advisable. To this end, the game brings a Professional list, with detailed profiles, to analyze their characteristics and determine which of them better fits the needs of the project at the time. Each professional has characteristics such as Teamwork, Leadership, Concentration, Technical Vision, Abstract Vision, Motivation, Hourly Cost, Productivity in Code, among others. These characteristics will influence the performance in the team project. It is up to the player dealing with different profiles so that he or she can maximize the results.

Due to the fact that the game is multiplayer, hiring or firing professionals require that the player remain attentive to other factors besides the professional characteristics. The list which includes the available professionals for recruitment, called market, is the same for all in the

Figure 2. Monitoring graphs

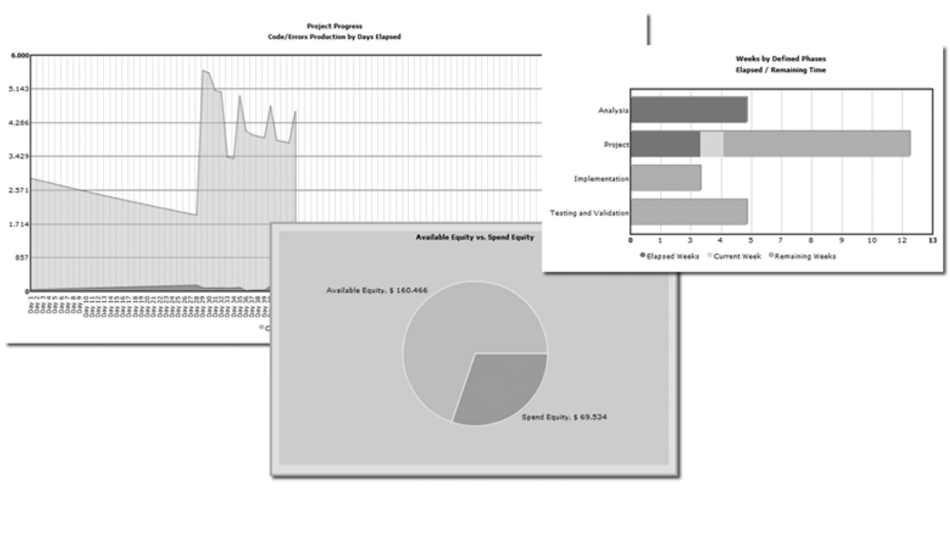
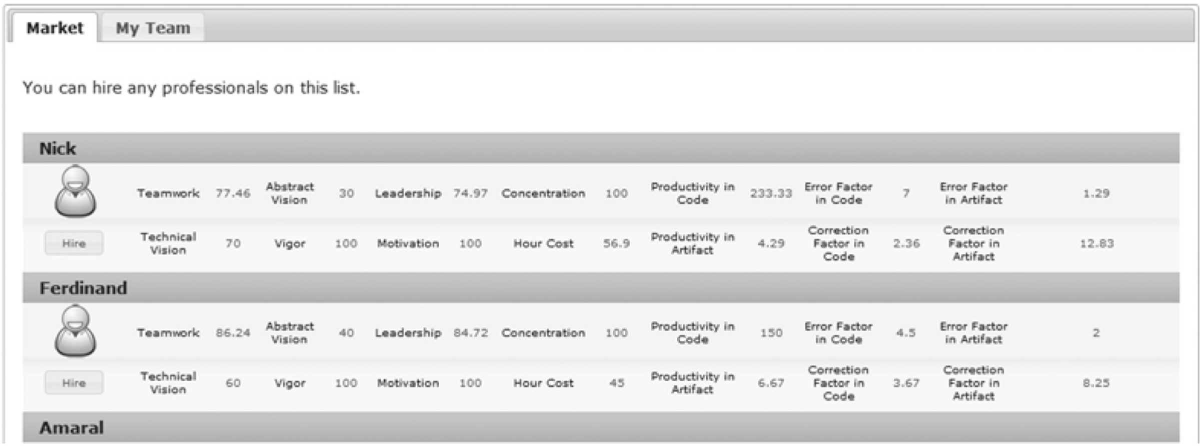


Figure 3. Professionals list



game. Thus the player should be cautious with the deadlines, because a professional that he or she intends to hire just for a specific phase may not be available anymore in that period. This same concern should also be watched over when firing, since in the game, the newly recruited professionals have a smaller productivity curve. That may hinder an exchange of these professionals in the final stages of a project phase.

The amount of available professionals on the market is controlled by the game in order to supply only the needs of ongoing projects. There is no excess of professionals, but it could have a shortage in certain moments.

To confer greater dynamics to the game, the professionals' attributes (Vigor, Motivation, etc) are not static. Over the days, professionals get tired, lose the motivation or even get distracted, affecting their productivity. To deal with this, the player can make new acquisitions, through which he or she can invest and revitalize the staff.

Acquisitions are material assets (printer, coffee maker, among others), courses, conferences (motivational or techniques) or case tools that improve the team productivity in some way. These tools include the overtime, which may be paid to increase the team activity. Other tools have preventive effect, for example, the acquisition of a backup system to avoid possible future problems with loss of data.

A new project, once started, brings fairly detailed planning done to the whole project and also for each stage: budget, time in weeks, artifacts and codes to be produced, among other details. Thereafter, the player's mission is to meet these targets and deadlines in the best possible way. However, it is given the option to negotiate the deadlines or budget. This fact entails a decrease in the project's final score.

Each project concluded, successful or not, gives the player a certain score, calculated in accordance with the fulfillment in a timely manner of each phase, financial balance or loss, production targets and quality affected. These points generate some rankings in which the player can measure his performance compared to other players.

The eRiskGame has its own time. The days elapse without the player's intervention and they progress differently from real time to prevent that the projects prolong for months or even years. This way, the player can prepare his strategy and let the game develop according to his planning, and may recur in his available schedule to verify the game's progress and to take other actions, if necessary.

This feature allows that several days take place without the need of the player to logon to the game. However, the player's absence for a prolonged time may bring the project to irreversible situations, since any inconvenience caused by agents may occur and if no action is taken in a timely manner, the progress of the project may be compromised.

Multiagent System

The focus of this serious game is on Risk Management. We noticed that one of the greatest difficulties of a project manager is to predict the problems that may arise and design plans to minimize their effects. A real project management environment is very complex to be represented, mainly due to the risks of the project that are involved in it, which affect the time schedule or resources, and due to the product risks, which affect software quality or performance. Business risks are also considered. They affect the organization that develops or purchase the software. To help representing this environment, intelligent agents were used to provide the game with some of these risks in different stages and different conditions for each project.

Before explaining a little bit about the logic behind some developed agents, it is convenient to expose which risks and consequences were modeled:

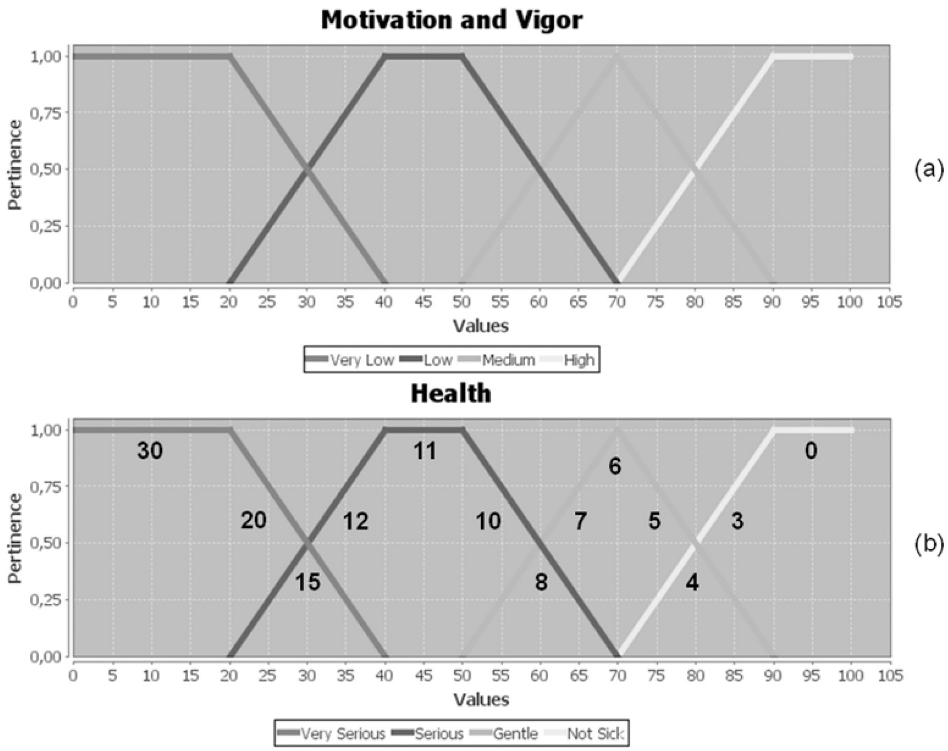
- Professional is ill, having to be away from work by a certain period, but maintaining his costs in the project;
- Valuable professional quits the team;
- Problems in the server resulting in partial or complete loss of the suitable code due to lack of backup system;
- Financial problems in the company forcing a reduction in the project budget;
- New similar software product launched on the market leading to a reduction within the deadline for completion;
- Inspection made in the software detecting several errors;
- Client requires new functionality not expected previously;
- Conflicts between professionals become routine and are starting to affect the productivity;
- Rumors in various aspects and professional relationships problems, affecting the staff concentration, motivation or performance.

Due to the diversity of problems, many agents had to be implemented, being each one of them in charge of monitoring several game aspects to act in their respective risks. However, agents are sociable and when one acts, providing some adverse situation; the others are informed, increasing or decreasing the risk of occurrence of other mishap. For instance, when a professional quits, the agent in charge for "rumors" may influence the other ones concentration.

Besides analyzing the professionals' attributes and the project progress, the agents that have been developed count with a peculiarity: their actions shall be subject to a probabilistic factor. All agents have a numeric attribute that represents the mishap (risk) occurrence probability. Then, when a state of the environment is detected, the action related to it will not always happen. A random number is chosen and compared with this attribute to determine whether the agent will or will not act. This way, we seek to provide a very common and necessary element in management environments: luck.

Some different AI techniques were used to provide intelligence to the agents, being the Fuzzy Logic[4] the most used. One of the applications that use this logic is the agent responsible for examining a professional's health conditions and keeps them off temporarily by medical leave. This is a simple reactive agent that monitors the motivation and the vigor (Figure 4-a) of all professionals who are designated to a project and identifies a health

Figure 4. Fuzzification and defuzzification graphs



condition, as shown in Figure 4-b, to determine how long this professional should be kept away.

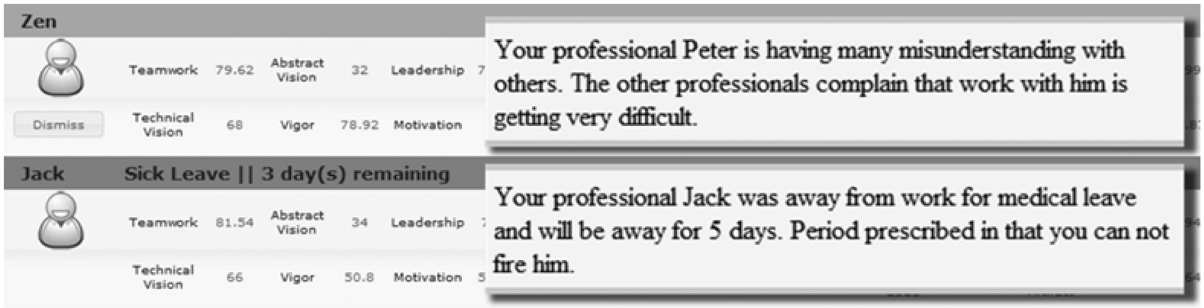
As the degradation of the attributes of the professionals is continuous, if there was not a factor of probability associated to the agents, as soon as a professional's health measured below 90, this one would already be kept off for 3 days, preventing that the physical state get worse until a most critical situation were identified. In this case, the associated probability is a very low number, which avoids the occurrence of deviations and allows that the number of days off be different in most cases.

The agent which creates rumors or discussions between the professionals is a model-based agent. It notes the quantity of professionals in each project, the time elapsed and the quantity of errors in the project, the leadership and team work of each professional for, from time to time, inform the players if any of them is clashing with the team or if the staff is concerned with any factor linked to possible changes in the company organization or even rumors involving the project discontinuity.

To detect certain states, the agent must inform the player about what has been occurred and stores data on this report. Subsequently, if the environment perceptions related to these data have degraded and no corrective action has been registered, the agent will create some situation in the project that will result in lack of concentration or high concern in the professionals, thereby increasing the risk of poor performance of several other agents.

The player is informed about the occurrence of any of these contingencies by means of messages in his inbox in addition to other reports located along the performance focus (Figure 5). The available message inbox is essential to maintain the player updated about the changes occurred in the project and it is one of the main communication tools included in the game.

Figure 5. Notices in the professional and inbox



Communication

The eRiskGame was designed on the Web platform, aiming at potentialization of the relations between students (players) and teachers (evaluators) through the use of a forum, chats and electronic messages. These communication means can be exploited by teachers in application and orientation of the activities developed in the game, besides sharing knowledge and acquired experience.

Students are no longer of passive receptors information in the classroom and become active participants in their learning environment (Xakaza-Kumalo, 2010). Thus, a teaching tool cannot be limited to transmit knowledge, without any feedback. It must allow students to exchange information and discuss about its content, so they could formulate questions and acquire new knowledge.

E-mail in the eRiskGame allows sending messages to only one receiver at a time. If the player wishes to send some information to a user group, he or she must create a favorite group and send the message to it.

As a synchronous communication way, the eRiskGame has a chat system to facilitate the exchange of instantaneous information among players. To access the system, the user will see a list of the existing chat rooms and their respective themes and will have the option of creating a new one and restrict the access with password, if desired. As soon as the user creator leaves the chat room, all participants will also be removed and the chat room shall be automatically closed.

A discussion forum was also developed to promote discussions and information exchanges among the participants. However, the forums' success depends on the number and quality of contributions. Without focus or concern about meeting certain standards, a productive and prosperous debate is impossible. Therefore, the topics and messages may be moderated[5] by any teacher.

These tools make the game more attractive, communicative and efficient. In addition, the communication is a foundation of the learning proposed by PBL, where the debate and the information exchange between the students favor the development of their own concepts and knowledge as well.

Evaluation Module

Education is a kind of characterized service with a high intangibility degree. Contrary to the assets' quality, which can be measured objectively through indicators such as durability and defect number, the educational service quality has a more abstract and intangible nature (Zeithaml et al., 1996). It may be difficult an assessment of its efficacy or even the adequacy of its use.

In these cases some measures may be taken to verify if the student is making proper use of the subjects exposed and if the tool actually works according to its purpose. It is possible to assess the quality perceived by the users, by questioning them if the tool reached their expectations and if they have acquired the worked content, or to evaluate the real knowledge acquired by attributing to the specialists (teachers) the role to analyze the results obtained by means of tests or other evaluation methods.

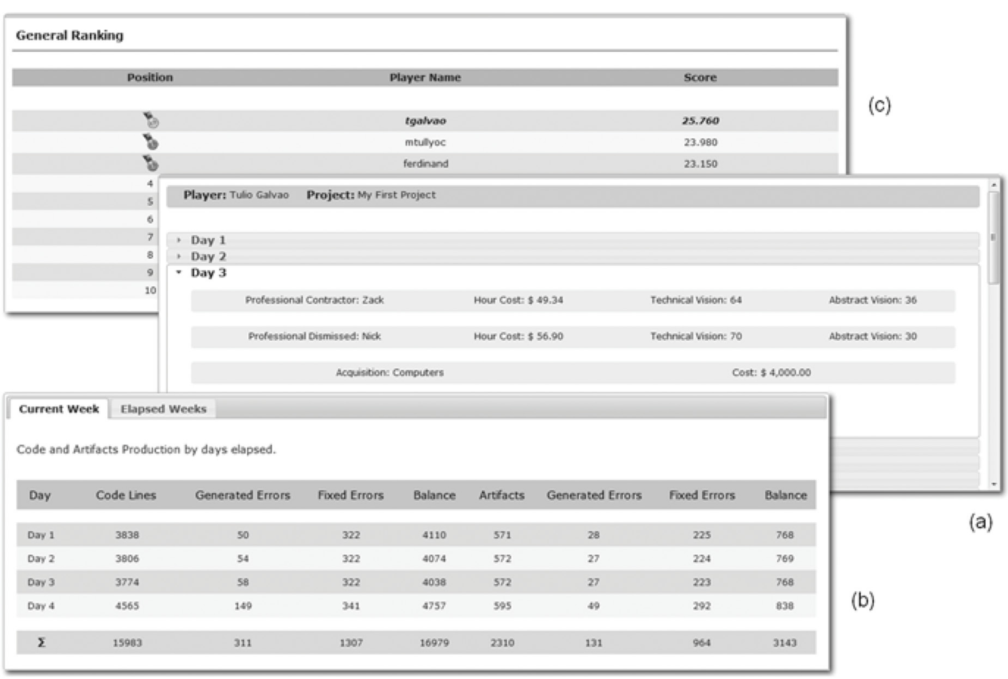
Having some kind of evaluation is important, because it allows the identification of deficiencies and the verification if the students are making progress in a wrong way. This way, teachers can guide when and how to perform in order to achieve better results. The eRiskGame is enabled with a special module designed to help the teachers in the task of monitoring the progress of the trainees and interpret their knowledge, skills and actions, in view of the expected changes in behavior, proposed in the tool goals.

Among these options available in that module, the registered teacher has access to:

- Monitoring the players' actions: all acquisitions, recruitment or any changes that the player makes in the project is recorded and shown in some kind of an interactive list of daily events (Figure 6-a);
- Monitoring the project is progress: the teacher has access to productivity statistics and projects' advancement (Figure 6-b);
- Rankings: the general ranking scores, project score and other classifications may also be accessed (Figure 6-c);
- Communication: the communication tool aggregated to the game can be accessed for guidelines, dialog and debates. The moderations in the forums may only be made by users which have joined through this module.

This set of alternatives allows the teachers to analyze the projects' progress and which actions have been taken at each moment, in addition to being able to compare the scores for a balance

Figure 6. Monitoring options



between the results of a certain player and the average of the others. The rankings provide a numerical comparison, but only through the actions history. The teacher is able to determine in which points the player may have taken some actions which jeopardized his good progress and guide him how to face certain obstacles ahead.

The assessment is an essential didactical task for the teaching work. By presenting a great complexity of factors, it cannot be summarized in simple tests execution and grades attribution. The measurement taken only gives quantitative data which must be appreciated qualitatively. The model used here intends to provide means to support the teachers' work as evaluators and explores their sensitivity to detect when and how to intervene in the learning process of their trainees.

Used Technologies and Utilization Purpose

The eRiskGame was implemented using technologies such as: PHP 5 as the server side scripting language, MySQL 5 as the database management system besides the intensive use of the JavaScript jQuery library, aiming to provide higher interactivity and usability to the game.

This tool may be used in software engineering formal courses (e.g. graduation courses, specialization courses) to train or even assess the abilities of the students in how to deal with situations and common decisions made by project managers. It will also be a good way for aspirants to job positions to exercise their abilities, at the same time that they exchange information and experiences with other players.

CONCLUSION

In this chapter a Persistent Browser-Based Game has been presented, the eRiskGame. This game was proposed as a tool to support the training of new project managers and was designed in line with the increasingly demanding users' needs, with different routines and schedules, motivating its development on the Web platform.

The methodology adopted has used the information technology potentialities allowing the creation of new teaching models. An efficient communication scheme was added to compose an environment where the teaching-learning process occurs in a more spontaneous way and involves the students more easily in the proposed context. In addition, the proposed tool offers a Risk Management practical vision different from the traditional education patterns.

Some earnings are easily noticed when using a game totally designed on the Web and without the need for special software to access it, such as: time flexibility, cost reducing, space flexibility and less interference in work routine. The high accessibility level promoted by this serious game also allows user to start a session in a given place and resume the game from any other place, using any access device. Thus, even people with very dynamic routines can benefit from the use of this tool.

Concerning the sharing and the quality of information exchange, one of the characteristics of the distance education systems, especially those which make use of asynchronous communication tools, such as forums, is that both the teachers and students have the opportunity to ripen their ideas and consult sources beforehand, favoring the preparation for more productive discussions. Thus the exchanged knowledge is better prepared and facilitates the users' understanding with different levels of familiarity with the theme for that matter.

All these benefits make this tool an innovative learning environment with characteristics very important to reach and gather distinct publics, allowing a greater collaboration between students and teachers, and could even be a model for other educational tools involving the most distinct topics.

REFERENCES

Ayas, K., & Zeniuk, N. (2003). Project-based learning: Building communities of reflective practitioners. *Management Learning*, *32*(1), 61–76. doi:10.1177/1350507601321005

Brewer, J. L. (2005). Project managers: Can we make them or just make them better? In *SIGITE '05, Proceedings of the 6th Conference on Information Technology Education* (pp. 167-173). New York, NY: Association for Computing Machinery.

Browser Games. (2010). *Browser games list*. Retrieved August 25, 2010, from http://www.browsergameslist.com

Chou, C., & Sun, C.-T. (1996). A computer-network-supported cooperative distance learning system for technical communication education. *IEEE Transactions on Professional Communication*, *39*(4), 205–214. doi:10.1109/47.544576

Dantas, A. R., Barros, M. O., & Werner, C. M. L. (2004). A simulation-based game for project management experimental learning. In *SEKE'04, International Conference on Software Engineering and Knowledge Engineering* (pp. 19-24). Banff, Canada: Conference on Software Engineering and Knowledge Engineering.

Drappa, A., & Ludewig, J. (2000). Simulation in software engineering training. In *ICSE '00, Proceedings of The 22nd International Conference on Software Engineering* (pp. 199-208). New York, NY: Association for Computing Machinery.

Frank, M. (2008, June). *Teaching technology management in a project-based learning environment*. Paper presented at the IEEE International Engineering Management Conference. IEMC Europe 2008, Estoril, Portugal.

Hsu, R. C., & Liu, W.-C. (2005). Project based learning as a pedagogical tool for embedded system education. In *ITRE 2005, 3rd International Conference on Information Technology: Research and Education* (pp. 362- 366). Hsinchu, Taiwan.

Navarro, E. O., & van der Hoek, A. (2004). SimSE: An interactive simulation game for software engineering education. In *CATE'04, Proceedings of the 7th IASTED International Conference on Computers and Advanced Technology in Education*. Kauai, Hawaii.

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003). Aspects of game-based learning. In *I-KNOW* 03, *Third International Conference on Knowledge Management* (pp. 217-224). Graz, Austria.

Pontes, A. A. A., Mendes Neto, F. M., & de Campos, G. A. L. (2010). Multiagent system for detecting passive students in problem-based learning. *Advances in Soft Computing*, *71*, 165–172. doi:10.1007/978-3-642-12433-4_20

Prensky, M. (2001). *Digital game-based learning*. New York, NY: McGraw Hill.

Project, P. (2009). *Persistent browser-based game – Defining a genre*. Retrieved December 18, 2010, from http://www.pbbg.org

Project Management Institute. (2004). *A guide to the project management body of knowledge* (3rd ed.). Newtown Square, PA: Project Management Institute.

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice Hall.

Sandford, R., Ulicsak, M. F. K., & Rudd, T. (2006). *Teaching with games: Using commercial off-the-shelf computer games in formal education*. Bristol, UK: Futurelab. Retrieved August 22, 2010, from http://www.futurelab.org.uk/projects/teaching-withgames/research/final-report

Schwalbe, K. (2002). *Information Technology: Project management*. Cambridge, MA: Course Technology.

Serious Games. (2010). *Serious games initiative*. Retrieved December 10, 2010, from http://www.seriousgames.org/

Sørensen, B., & Meyer, B. (2007, September). *Serious games in language learning and teaching: A theoretical perspective*. Paper presented at DiGRA Conference, Tokyo, Japan.

Susi, T., Johannesson, M., & Backlund, P. (2007). *Serious games - An overview*. Technical Report HS-IKI-TR-07-001, School of Humanities and Informatics, University of Skovde, Sweden.

Xakaza-Kumalo, S. (2010). Using the internet communication tools to facilitate learning. In ICEMT 2010, *International Conference on Education and Management Technology* (pp. 706-710). Cairo, Egypt.

Yadav, S. S., & Xiahou, J. (2010). Integrated project based learning in software engineering education. In *ICENT 2010, International Conference on Educational and Network Technology* (pp. 34-36). Qinhuangdao, China.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31–46. doi:10.2307/1251929

Zynga Inc. (2011). *FarmVille: Games in Facebook*. Retrieved February 10, 2011, from http://www.facebook.com

KEY TERMS AND DEFINITIONS

Game-Based Training: Use of games designed with educational emphasis in training or specific skill qualification.

Intelligent Agents: Artificial Intelligence technique that uses an entity of software or hardware to perceive and perform in a certain environment.

Persistent Browser-Based Game: Computer Game entirely implemented in a browser that does not use any special software to be activated and keep the session data of the

Project Management: Procedure whereby a venture is led to its conclusion taking into account the costs, deadlines, quality expectation and other targets set in advance.

Project-Based Learning: Use of projects which represent a real situation to facilitate the learning and to evaluate the student competences.

Risk Management: Activity linked to the Management Projects which deals with the monitoring and eventual combat to the risks involved in a software project.

Software Engineering: Computer knowledge area oriented for specification, development and software maintenance systems.

ENDNOTES

1. Allow the participation of only one player by match.
2. Allow many players to participate simultaneously of the same match.
3. Generally refers to delineate, by a computer application, of the best route between two points.
4. The Fuzzy Logic can be seen as a research area on the treatment of uncertainty. Is an extension of the Boolean Logic that allows intermediary logic values between the FALSE (0) and the TRUE (1).
5. It is known as any device moderator, human or not, which allows supervise and suppress the messages not complying with the letter for use of the forum or which can cause legal procedures.

This work was previously published in Handbook of Research on Serious Games as Educational, Business and Research Tools, edited by Maria Manuela Cruz-Cunha, pp. 232-249, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 68
Effect of Customer Power on Supply Chain Integration and Performance[1]

Xiande Zhao
Chinese University of Hong Kong, Hong Kong

Baofeng Huo
Xi'an Jiaotong University, China

Barbara B. Flynn
Indiana University, USA

Jeff Hoi Yan Yeung
Chinese University of Hong Kong, Hong Kong

ABSTRACT

Supply chain integration (SCI) has received increasing attention from academic researchers and practitioners in recent years, however, our knowledge of what influences SCI, and how SCI influences the performance of supply chains and manufacturers within the supply chain is still very limited. Although researchers in marketing and management have investigated power and relationship commitment issues within and between organizations, few have examined their impact on SCI. This chapter studies the relationship between power, relationship commitment and integration between manufacturers and their customers within a supply chain. The impact of customers' SCI on the customer service and financial performance of manufacturers is also investigated. The authors propose and empirically test a model using data collected from 617 manufacturing companies in China. The results show that customers' use of different types of power has different impacts on manufacturers' relationship commitment. Expert power, referent power and reward power are very important in improving manufacturers' normative relationship commitment, while reward power and coercive power enhance instrumental relationship commitment. The authors also find that normative relationship commitment have a greater impact on customer integration and customer service by manufacturers than instrumental relationship commitment. Customer integration significantly enhanced manufacturers' customer service and financial performance. The improvement in customer service of manufacturers positively influenced their financial performance.

DOI: 10.4018/978-1-4666-1945-6.ch068

INTRODUCTION

Global competition and escalating customer expectations have forced manufacturing companies to focus more on delivery speed, dependability and flexibility to meet changing customer requirements (Boyer & Lewis, 2002; Flynn & Flynn, 2004). To enhance these capabilities, many companies have implemented supply chain integration (SCI) strategies (Bowersox, Closs & Stank, 1999). SCI is defined as the degree to which a firm strategically collaborates with its supply chain partners and collaboratively manages intra- and inter-organization processes, to achieve effective and efficient flows of products, services, information, money, and decisions, with the objective of providing maximum value to customers at low cost and high speed (Bowersox, et al., 1999, Frohlich & Westbrook, 2001). Extensive literature has cited the importance of SCI in achieving a competitive advantage (Bowersox & Morash, 1989; Lee & Billington, 1992, Morris & Calantone, 1991), as well as in enhancing operational performance (Ahmad & Schroeder, 2001; Frohlich & Westbrook, 2001; Johnson, 1999; Narasimhan & Jayaram, 1998; Stank, Keller & Closs, 2001a). However, limited empirical research has been done on the antecedents of SCI. Stank, et al. (2001) classified SCI into six types: internal, customer, material/service supplier, technology and planning integration, measurement, and relationship integration, developing and testing an instrument for measuring the effect of SCI competencies in different types of integration on firm performance, finding that customer integration was the dominant competence influencing overall and individual performance. Because of the importance of customer integration in improving a firm's performance, this chapter focuses on the integration between manufacturers and their customers.

Although researchers and practitioners have realized the importance of SCI, our understanding of what enables SCI is still very limited. Cox (2001) addressed the importance of power in business strategy and operational performance, finding that competence in supply chain management should start from understanding the power and business strategy of suppliers. He indicated that understanding the supplier's power perspective could enhance effective procurement and supply management, introducing a power matrix which illustrated how a buyer can enhance its competence. However, this study was limited to the author's personal observations, with no empirical evidence.

Maloni and Benton (2000) examined the impact of different types of power on the strength of relationship between auto parts suppliers and automobile manufacturers and how it influenced the performance of supplier, manufacturers and the supply chain. Relationship strength was operationalized by five elements: commitment, conflict, conflict resolution, cooperation, and trust. They found that expert and referent power was positively related to the strength of the relationship, while coercive and legal legitimate power were negatively related to it. They also found that a stronger relationship was related to the performance of suppliers, customers and the supply chain. In another study, Benton and Maloni (2005) investigated the relationship between: 1) the use of different types of power by the supplier, 2) strength of the supplier-manufacturer relationship, 3) performance of the supplier, manufacturer, and the supply chain, and 4) supplier satisfaction. The power-affected manufacturer–supplier relationship was positively related to supplier performance and satisfaction, however, they failed to find a causal relationship between performance and satisfaction.

While there is a dearth of research on the factors that influence SCI, marketing researchers have studied factors that influence inter-firm relationships in general. Brown, Lusch and Nicholson (1995) empirically investigated the impact of power and relationship commitment on marketing channel members' performance from a relationship marketing perspective. They found that suppliers' use of power was related to retailers' commitment to the channel relationship, which was associated with the perceived performance of the supplier and subsequent financial performance of the retailer. They demonstrated the moderating effect of power asymmetry within the channel, finding that suppliers' use of non-mediated bases of power was related to retailers' normative relationship commitment, but negatively associated with their instrumental relationship commitment. In contrast, suppliers' use of mediated power negatively was related to both retailers' normative and instrumental relationship commitment. They also found that enhancement in normative relationship commitment was related to improved performance of suppliers and financial performance of retailers. However, higher instrumental relationship commitment was negatively associated normative relationship commitment and, subsequently, had a detrimental effect on performance for all channel members.

Thus, based on our review of the literature on inter-firm relationships, we found that customer integration is very important in improving firm performance. Furthermore, power and relationship commitment are important factors in inter-firm collaboration. However, to the best of our knowledge, there are no solid empirical studies that investigate the relationship between the use of power, relationship commitment, customer integration and the performance of manufacturers within a supply chain. In this chapter, we investigate this relationship. Specifically, our objectives are:

1. To develop and test a measurement instrument for power, relationship commitment, customer integration and manufacturing performance within a supply chain.
2. To propose and empirically test a model that represents the relationship among power, relationship commitment, customer integration and manufacturing performance in a supply chain.
3. To offer guidelines for practicing managers to enhance their performance through integration and better management of customer relationships.

We based our study in China, which has become a very important manufacturing base in the world. The average annual growth of GDP in China has averaged around 10% for the last 15 years, with the manufacturing component of GDP growing at about 40% annually; thus, manufacturing drives the Chinese economy. In order for Chinese manufacturers to compete in the global marketplace, they must strive to improve their supply chain operations, and customer integration is an important of enhancing global competitiveness. China's dynamic competitive environment provides a fertile ground for investigating power, relationship commitment and their impact on SCI. Since most previous studies of power and relationship commitment were conducted in the U.S., basing our research in China allows us to examine whether findings from previous studies can be applied in a rapidly developing economy, with a non-Western national culture.

The chapter is organized as follows: we first address the theoretical background of the study with a conceptual model, proposing associated hypotheses. This is followed by discussion of the research methodology, analysis and results. Finally we summarize our major findings and present the conclusions and limitations of the study.

BACKGROUND

Power and Relationship Commitment

Marketing researchers have investigated the impact of power on cooperation between members of distribution channels (Brown, et al., 1983, Brown, et al., 1995, Goodman & Dion, 2001). Channel power is defined as the ability of one channel member to influence the decisions of another channel member (Brown, et al., 1983; Brown, et al., 1995, Goodman & Dion, 2001). Studies of channel power draw upon the seminal work of French and Raven (1959), who classified power into five types, summarized in Figure 1. They can be further classified into mediated and non-mediated sources of power (Brown, et al., 1995, Maloni & Benton, 2000; Benton & Maloni, 2005), reflecting "…. whether the source does or does not control the reinforcement which guides the target's behavior" (Tedeschi, Schlenker & Lindskold, 1972, p.292). Reward and coercive power are considered mediated because reinforcement is controlled by the customer. The customer, as the source of the power, decides whether, when and how to use its power to influence the manufacturer's (target's) decision and behavior. Expert, referent and legitimate power are considered non-mediated, because the manufacturer, itself, decides whether and how much it will be influenced by them. Non-mediated power is more relational and positive in orientation (Brown, et al., 1995, Maloni & Benton, 2000), while mediated power represents the competitive and negative uses of power traditionally associated with organizational theory (Brown, et al., 1995, Maloni & Benton, 2000).

Another very important concept in inter-organizational relationships is relationship commitment, which is the willingness of a party to invest resources in a relationship (Morgan & Hunt, 1994). Morgan and Hunt (1994) suggest that the propensity for relational continuity and the establishment of a long-term relationship are the primary themes of relationship commitment. It can be classified as normative or instrumental (Brown, et al., 1995). Normative relationship commitment is the willingness of an organization to secure a relationship, due to its identification with and emotional attachment to the goals and values of another party (Morgan & Hunt, 1994; Wetzels, de Ruyter & van Birgelen, 1998). It is intrinsic, based on identification and internalization of the norms and values that are held in common (Brown, et al., 1995). Identification occurs when a manufacturer accepts a customer's influence because it admires the way that the customer manages its business and thus wants to establish or maintain a satisfying relationship with it, and internalization occurs when a manufacturer accepts a customer's influence because it holds values and norms of behavior that are similar to the customer. In contrast, instrumental relationship commitment is based on compliance (Brown, et al., 1995), which occurs when one party accepts influence from another, in the hope of receiving a favorable reaction from the other party. It is driven by the potential for extrinsic rewards or punishment. If a manufac-

Figure 1. Bases of inter-firm power

Power Base	Description	Example of Customer's Use of this Type of Power
Expert power	Source has knowledge, expertise or skills desired by the target	The manufacturer's customer knows what the final customer wants or has knowledge and expertise in designing or distributing new products to the final consumers
Referent power	Target values identification with the source	The manufacturer desires association with the customer because it is well known and admired by the manufacturer. If the customer has developed a strong bond through its demonstrated concern, management style and organizational personality, the customer has power over the manufacturer, based on positive emotional ties (Goodman and Dion 2001)
Legitimate power	Target believes source retains natural rights to influence.	The manufacturer believes that the customer has the right to request and expect things to be done according to its requirements, as part of the supplier and customer relationship. This is a result of the level of importance accorded the customer in the supply chain.
Reward Power	Source retains ability to mediate rewards to target	The customer has the ability to provide rewards that are attractive to the manufacturer, for example, the customer can decide to give more business to the manufacturer.
Coercive power	Source holds ability to mediate punishment to target	The customer has the ability to provide punishments that are detrimental to the manufacturer, for example, the customer can cancel business or reduce the volume of business with the manufacturer.

Adopted from: Maloni and Benton (2000). "Power influences in the supply chain," *Journal of Business Logistics*, Vol. 21, No. 1 2000, p. 54.

turer's willingness to invest in and maintain a relationship is based on calculation of benefits and cost and is driven by rewards and punishment offered by the customer, the relationship commitment is instrumental.

The use of non-mediated power was found to enhance positive attitudes towards channel relationships (Frazier & Summers, 1986) and foster congruence in values and norms between channel members. Therefore, the use of expert, referent and legitimate power leads to a higher degree of normative commitment between partners (Brown, et al., 1995). At the same time, the more one party uses non-mediated power to influence the other party, the more it focuses on common norms and values, as well as on the relationship itself. "As these intrinsic factors become central, extrinsic factors such as rewards and punishments, become less important" (Brown, et al., 1995, p.368). Therefore the use of non-mediated power by the customer decreases the degree of instrumental relationship commitment by the manufacturer. Generalizing this literature to supply chains, we propose the following hypotheses (see Figure 2):

H1a: *Greater expert power by the customer will be associated with stronger normative relationship commitment by the manufacturer.*

H1b: *Greater referent power by the customer will be associated with stronger normative relationship commitment by the manufacturer.*

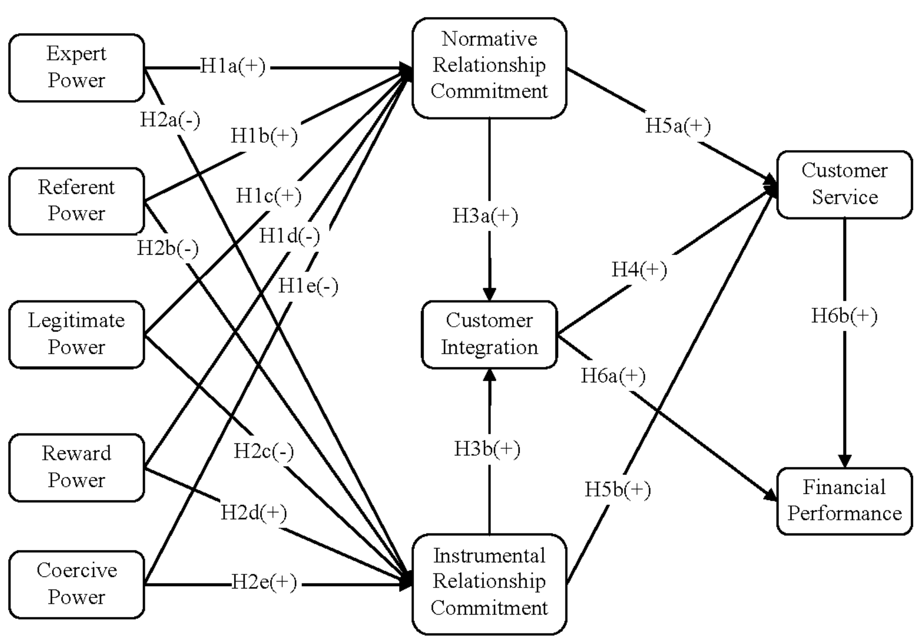
H1c: *Greater legitimate power by the customer will be associated with stronger normative relationship commitment by the manufacturer.*

H2a: *Greater expert power by the customer will be associated with weaker instrumental relationship commitment by the manufacturer.*

H2b: *Greater referent power by the customer will be associated with weaker instrumental relationship commitment by the manufacturer.*

H2c: *Greater legitimate power by the customer will be associated with weaker instrumental relationship commitment by the manufacturer.*

Figure 2. Proposed model



When a customer uses reward and coercive power, it provides extrinsic motivation for the manufacturer's commitment to it, and the manufacturer is driven to comply with the customer's requirements, in order to achieve favorable outcomes. Therefore, we expect that the customer's use of mediated power will foster stronger instrumental relationship commitment by the manufacturer (Kasulis & Spekman, 1980; Brown, et al., 1995). However, the frequent use of mediated power can damage relational norms (Boyle et al., 1997) and cooperation (Skinner, Gassenheimer & Kelley 1992), reducing the strength of the relationship (Benton & Maloni, 2005; Maloni & Benton, 2000). Therefore, we expect that the customer's use of mediated power will decrease normative relationship commitment. Based on the above discussion, we propose the following hypotheses:

H1d: *The perception of greater customer reward power is inversely related to normative relationship commitment by the manufacturer.*

H1e: *The perception of greater customer coercive power is inversely related to normative relationship commitment by the manufacturer.*

H2d: *The perception of greater customer reward power is directly related to instrumental relationship commitment by the manufacturer.*

H2e: *The perception of greater customer coercive power is directly related to instrumental relationship commitment by the manufacturer.*

Impact of Relationship Commitment on SCI

The need for integration in the physical distribution of products has been articulated by researchers since the 1970s, when Lambert, Robeson and Stock (1978) considered distribution activities as a process to be managed. However, not until a decade ago did researchers start to call for a systems approach to supply chain integration (SCI). The purpose of SCI is to integrate the relationships, activities, functions, processes and locations among all channel members of a supply chain (Bowersox & Morash, 1989; Hammer, 1990). This primarily involves planning, coordinating and controlling materials, parts and finished goods from suppliers to customers at the strategic, tactical and operational levels (Stevens, 1989). The goal of integration is to eliminate all boundaries, in order to smooth the flow of material, cash, resources and information (Naylor, Nam & Berry, 1999). Based on this line of thought, SCI should be strategically managed as a single system, as opposed to individually optimizing fragmented subsystems (Vickery et al., 2003).

Stank, et al. (2001a) developed and tested an instrument for measuring SCI competencies, as well as evaluating their relative importance in developing logistic distinctiveness. Among the six areas of SCI they identified, customer and internal integration were the most significant contributors to overall performance. Johnson (1999) developed and empirically validated strategic integration as a center of gravity for building up sustainable integration. His findings revealed that dependence, flexibility, continuity expectations, and relationship age have a positive effect on a distributor's strategic integration with its suppliers, which enhances the distributor's financial performance. Frohlich and Westbrook (2001) investigated supplier and customer integration strategies, in developing scales for measuring SCI, and five types of integration (inward-, periphery-, supplier-, customer-, and outward-facing) were identified in the sample.

Among the types of integration identified in the literature, customer integration has been found to have the most significant impact on performance (Stank, et al., 2001a). It is defined as a core competence derived from better coordination of all critical customers in a supply chain, in order to jointly achieve improved service capabilities, at lower total supply chain cost (Bowersox, et al., 1999). In order to have a higher degree of customer integration, there has to be information sharing, coordination and synchronization of processes between a manufacturer and its customers. Therefore, SCI is created by cooperative, mutually beneficial partnerships between supply chain members (Wisner & Tan, 2000); thus, relationship commitment plays a very important role in SCI.

Morgan and Hunt (1994) proposed and tested a number of hypotheses concerning the relationship between relationship commitment, trust, their precursors and their outcomes. They found that relationship commitment was positively related to acquiescence and cooperation in a relationship, but negatively related to propensity to leave. When there is relationship commitment, supply chain partners become integrated into their key customers' business processes and more tied to established goals (Chen & Paulaj, 2004). Since integration between manufacturers and customers in a supply chain requires an investment of time, human resources and financial capital, return on the investment in integration may not be realized immediately, and normative relationship commitment may be required. If a manufacturer normatively commits to its customer, it will want to operate its company as the customer does, cooperating with it. Thus, normative relationship commitment can improve the extent of customer integration. In order to avoid loss or punishment or obtain rewards from the customer, a manufacturer will integrate with its customer's activities, in order to serve it better. Thus, although instrumental relationship commitment can help with SCI, its impact will be much lower. Therefore, we propose the following research hypotheses:

H3a: *Normative relationship commitment by the manufacturer will be associated with the degree of the integration between a manufacturer and its customer.*

H3b: *Instrumental relationship commitment by the manufacturer will be associated with the degree of the integration between a manufacturer and its customer.*

H3c: *Normative relationship commitment by the manufacturer will be associated with the degree of integration than instrumental relationship commitment.*

Impact of Relationship Commitment and Customer Integration on Performance

A common measure of business performance is financial performance, since the primary goal of a business is to make profits for its shareholders (Chen & Paulraj, 2004). Financial performance can be evaluated on many dimensions (Bender, 1986; Boyer et al., 1997; Boyer, 1999), however, much of the literature (e.g., Dixon, Nanni & Vollmann, 1990; Eccles & Pyburn, 1992; Hall, 1983; Johnson & Kaplan, 1987; Skinner, 1971) pinpoints the limitations of relying solely on financial performance measures. Van Hoek (1998) advocated that supply chain firms devise innovative measurement systems, as opposed to traditional ROI-based systems. For example, Vickery, et al. (2003) included the dimension of service performance in their customer service construct, including pre-sale customer service, product support, and responsiveness to customers, delivery dependability and delivery speed. Supply chain studies (Tan, Kannan & Handfield, 1998; Frohlich & Westbrook, 2001; Vickery, et al., 2003) have used both operational and financial performance as indicators of organizational performance. In this chapter, we included both operational and financial performance of the manufacturer as performance measures.

It has been widely accepted that operations priorities can be expressed by at least four dimensions: cost, quality, delivery, and flexibility (Ward, Duray, Deone & Sun, 1995; Kathuria, 2000; Dangayach & Deshmukh, 2001; Christiansen, Berry, Bruun & Ward, 2003). White (1996) suggested that performance measures derive from operations priorities, including cost, quality, flexibility, delivery reliability and speed. The last two types of performance can be considered delivery performance. Just as Aggarwal (1997) described the seventies as "the decade of productivity", the eighties focused on total quality management," and the nineties belong to flexibility. While cost and quality have been emphasized a great deal by companies in the past twenty years, flexibility has become one of the most important operations priorities in recent years, and on-time, reliable and high-speed delivery are being required by more and more customers. Thus, flexibility, delivery, and customer service are important competitive priories in today's global competitive environment, and SCI is designed to enhance capabilities in all three areas (Davis, Aquilano & Chase, 2003). The performance measures used in this study are customer service related measures, including delivery and flexibility performance.

Several studies have examined the relationship between SCI and performance (Armistead & Mapes, 1993; Daugherty, Ellinger & Gustin, 1996; Tan et al., 1998). Armstead and Mapes (1993) found that greater integration of the supply chain led to improved operating performance. Stank, Keller and Daugherty (2001b) found that collaboration with supply chain partners facilitated internal collaboration, which in turn enhanced logistics performance. Stank, et al. (2001b) described customer integration as the most significant factor in influencing overall and individual performance. Frohlich and Westbrook (2001) demonstrated that supply chain companies with the widest degree of the arcs of integration achieved the highest level of improvement in customer service, on-time delivery, and delivery lead time. Vickery et al. (2003) concluded that SCI can improve service performance. Benton and Maloni (2005) found that the strength of the relationship between suppliers and automobile manufacturers had a positive impact on the performance of suppliers, manufacturers, and the supply chain.

Process integration between manufacturers and customers helps to simplify procedures and synchronize boundary-spanning activities. As a result, manufacturers achieve faster and more dependable deliveries. Information sharing and coordination of activities between organizations further enhances the manufacturers' ability to quickly introduce new products or modify existing products to meet customers' needs. In addition, manufacturers can respond more quickly to changes in demand volume and the mix of products needed in the marketplace. As a result, customer service is improved. Therefore we propose the following hypothesis:

H4: *The extent of customer integration will be associated with manufacturers' customer service.*

In addition to better integration with its customers, relationship commitment can also enhance manufacturers' performance in other ways. Brown, et al. (1995) found that normative relationship commitment was related to the retailer's perception of supplier performance, which, in turn, was associated with the retailer's performance. However, they also found that the retailer's instrumental relationship commitment negatively related to perceived performance of the supplier. When there is a higher level of normative relationship commitment, the manufacturer identifies with and internalizes with the value and norms of the customer. Therefore, they have congruence in their values, management philosophies and norms of behavior, leading to better cooperation and reducing the potential for conflict. As a result, normative relationship commitment is associated with the manufacturers' operational performance. However, when relationship commitment is instrumental, the two parties do not share the same values and norms of behavior, and each is looking for short-term rewards. The result is that there is less cooperation and a greater chance for conflict. Therefore, instrumental relationship commitment will not be strongly related to customer service.

Based on the above discussion, we propose the following hypotheses:

H5a*: Normative relationship commitment will be associated with manufacturers' customer service performance.*

H5b: *Instrumental relationship commitment will be associated with manufacturers' customer service performance.*

H5c: *Normative relationship commitment will be more strongly related to customer service performance than instrumental relationship commitment will.*

Evaluation of manufacturers' financial performance uses financial and market measures to evaluate efficiency and effectiveness, providing a common measure of business performance (Chen & Paulraj, 2004). Chang and Chen (1998) found that there is a relationship between service quality and profitability. Vickery, et al. (1997) showed that delivery performance was related to several different measures of business performance, including ROI, ROI growth, ROS growth, and sales growth. Vickery et al. (2003) found that customer service performance was positively related to financial performance. With customer integration, manufacturers and customers are more likely to share information about forecasting, manufacturing, and inventory with each other. As a result, manufacturers can develop better manufacturing plans, reduce waste and maintain lower inventory levels, reducing inventory costs. Thus, customer integration is associated with manufacturers' financial performance through reducing costs. Customer service performance has a positive influence on financial performance through increasing revenues. If a manufacturer has a high level of customer service performance, it will meet the customers' timeliness and flexibility requirements, and satisfied customers will buy more products or bring in new customers. As a result, the manufacturer's sales are enhanced, market share is improved, and strong financial performance is achieved. In order to examine the importance of a manufacturer's customer service performance in improving its financial performance, we test the following hypotheses:

H6a: *The extent of a manufacturer's customer integration will be associated with its financial performance.*

H6b: *The customer service performance of a manufacturer will be associated with its financial performance.*

The proposed hypotheses are summarized in the model shown in Figure 2. The key constructs and the associated measurements items are shown in Appendix.

METHODOLOGY

Sampling and Data Collection

We collected data from manufacturing companies in China. Since China is becoming a strong global manufacturing base, its manufacturers play an important role in many supply chains. However, little solid empirical research has been done about supply chain management issues using data from China. Therefore, the results from this study can be of great value.

Random sampling was used to collect the data. Since China is very large, we strategically chose five cities, to provide geographic and economic diversity. Chongqing is a traditional industrial base, located in the central part of China. It is at a relatively lower stage of economic reform. Tianjin is an industrial city in northern China and reflects the "average" stage of economic reform and marketization in China. Guangzhou and Shanghai are in southern and eastern China and have enjoyed a higher degree of economic reform.

We also included Hong Kong, because Hong Kong is one of the freest economies in the world, where manufacturers operate in a quite different environment than the other cities of China. Most Hong Kong manufacturers have their manufacturing facilities in mainland China and thus are directly connected to the mainland economy. We believe that this sample provides a good representation of manufacturing activities in China.

To get a representative sample of manufacturing companies in the mainland China cities, we used the Yellow Pages of China Telecom. In Hong Kong, we used the directory of the Chinese Manufacturers Association. One important challenge was how to collect reliable data. After visiting fifteen companies in the pilot study, we determined that the best approach was to designate a key informant in each company who was knowledgeable about supply chain management. Key informants included the supply chain manager, CEO/president and vice president in charge of sales/marketing. Many studies have used a single informant in studying relationship issues between different organizations (e.g., Hewett & Bearden, 2001), however, some have demonstrated that using multiple informants is better (Bruggen, Van Lilien & Kacker, 2002). Due to time and budget constraints, we used single informants in this study.

We made telephone calls to randomly selected companies to find out whether they were manufacturers and identify the contact information of the key informants. We sent the questionnaire to them, along with a cover letter highlighting the objectives and potential contributions of the study. Respondents were encouraged to participate through the promise of a summary report of this study and a small participation incentive gift. Self addressed, stamped envelopes were sent along with the survey, to facilitate the return of the completed questionnaires, and follow-up telephone calls were made to improve the response rate. If there was excessive missing data, respondents were contacted by phone for clarification. Out of the 4,569 companies contacted, a total of 1,356 agreed to receive the questionnaire. After several follow-up calls, 617 usable questionnaires were received. The response rate, based on the number of companies contacted via telephone, was 13.5%, however, it was 45.5% based on the number of questionnaire distributed, comparable or better than the response rates of previous similar studies (Stank, et al., 2001; Frohlich & Westbrook, 2001).

Questionnaire Design

To develop reliable and valid measures of the constructs, we first went through intensive study of the literature to identify valid measures for related constructs. Whenever possible, we adopted measurement items that had been used and tested in previous studies. When the measurements items had not been well documented and tested in the literature, we proposed new measurement items, based on our understanding of the constructs, our observations during company visits and interviews with practitioners knowledgeable about supply chain management.

The measures for expert, referent, legitimate, reward and coercive power were adopted from Brown, et al. (1995). The measures for legitimate power were all related to the natural right of a customer to influence a manufacturer, and no items were designed to measure power based on judicial or legal right. Through our interviews, we found that legal legitimate power was not a big concern for customers, perhaps because the Chinese economy is still developing, and regulations for economic activities are not well formed, with few companies using legal instruments for supplier management. Thus, only traditional legitimate power was used in our study. Respondents were asked to indicate the degree of their agreement with statements concerning their major customer's power, using a Likert scale where "1" indicates "strongly disagree" and "7" indicates "strongly agree." The measures for relationship commitment were also adopted from Brown, et al. (1995). We used a total of six items to measure normative relationship commitment, and instrumental relationship commitment was measured using three measurement items.

The measures for customer integration were selected from items used by Narasimhan and Kim (2002), and Frohlich and Westbrook (2001). We selected those that were most relevant to information sharing and collaboration activities, based on observations during our plant visits and in-depth interviews with executives. Respondents were asked to indicate the extent of integration or information sharing between their organization and their major customer in 11 areas using a Likert scale with "1" being "Not at all", and "7" being "Extensive." Customer service, including flexibility, delivery and level of customer service, was measured by six items. Flexibility was operationalized as a manufacturer's ability to quickly modify products to meet its major customer's requirements, to introduce new products quickly and to respond to changes in market demand quickly, modified from items in several previous studies (Beamo, 1999; Frohlich & Westbrook, 2001). Delivery performance was operationalized as on-time delivery and lead time, and customer service was operationalized as overall customer service to the major customer, adapted from previous studies (Frohlick & Westbrook 2001; Stank, et al., 2001). Financial performance was measured by five items partly adopted from the measurement items in Narasimhan and Kim (2002), which measure the manufacturer's growth in sales, profit, market share, ROI and ROS. These indicators were measured using a Likert scale, with "1" for "much worse", "7" for "much better" than the primary competitor. The questionnaire also included items on the demographic profile of the company. The measurement items are listed in the Appendix.

In order to ensure reliability, an English version of the questionnaire was developed, then translated into Chinese by an operations management professor in China. The Chinese version was then translated back into English by a different operations management professor in Hong Kong, and the translated English version was checked against the original English version for question accuracy. In mainland China, the Chinese version of the questionnaire was used, while, in Hong Kong, the bilingual version was used. Before launching the full-scale study, the questionnaire was pilot tested using a sample of fifteen companies. We visited these companies and conducted face-to-face discussions with executives after they completed the questionnaire. Based on their feedback, we modified the wording of some questions and added or deleted some questions, making our questions more understandable and more relevant to practices in China.

Since we used a single informant to answer the self-reported questions in this study, we checked for potential common method bias. The items comprising the power, relationship commitment, customer integration and performance scales were not highly similar in content and the constructs were measured through 3-11 items. We used Harman's one-factor test of common method bias on the SCI and performance variables (Podsakoff & Organ, 1986; Podsakoff, et al., 2003; Hochwarter, et al., 2004), finding that several distinct factors were obtained among all the variables. Thus, we concluded that common method variance bias was not a problem.

ANALYSIS AND RESULTS

Respondent Profiles

The companies surveyed represent a large variety of the companies from a number of industries. More than a quarter of the companies were from metal, mechanical and engineering, followed by 17.86% in the textiles and apparel industry and 13.15% in electronics and electrical companies. Over 32% of the responding companies had annual sales of less than HK$5 million, while 14.99% had annual sales of HK$100 million or more. 40% of the respondents were top managers.

57% were functional managers and 30% were from marketing. 27% had been in their positions for one to three years, and nearly half had taken their positions in the past 6 years. About a fourth of respondents had been in charge of their department or company for 7 to 12 years, while another quarter had been in their positions for at least 12 years. Three quarters of the respondents had been in their position for more than 3 years. Thus, the respondents were familiar with the companies' activities, and the data collected from them should be reliable.

Measurement Development

A rigorous instrument development process was used to develop and validate the instrument, to ensure content validity, unidimensionality, reliability, convergent and discriminant validity. This procedure was adopted from previous empirical studies in logistics and supply chain management (Chen & Paulraj, 2004; Min & Mentzer, 2004; Garver & Mentzer, 1999). Prior to data collection, the content validity of the instrument was supported by previous literature and pilot tests. After data collection, we performed a series of analyses to test the reliability and validity of the constructs.

Unidimensionality and Reliability

Because few studies about SCI have been conducted in China, most of the measurement items were borrowed from previous research conducted in Western countries. A strict process for scale development was followed, particularly since the scales were being used in a very different national culture. We followed the two-step method used in Narasimhan and Jayaram (1998) to test construct reliability, employing exploratory factor analysis (EFA) to ensure unidimensionality of the scales, then Cronbach's alpha for assessing construct reliability. EFA was used with principal components analysis for data reduction and determining the main constructs measured by the items. The commonly recommended method of varimax rotation with Kaiser normalization was used to clarify the factors (Loehlin, 1998). Some measurement items were dropped, after comparing their loading on the construct that they were intended to measure to their loadings on other constructs. Through EFA, we tested the unidimensionality of the measurement scales. Cronbach's alpha was then computed for each construct, to test for internal consistency. Using the intercorrelation matrix, items with a correlation value below the 0.30 cutoff value were discarded (Flynn, Schroeder & Sakakibara, 1994). These steps were performed iteratively, to determine the items to be deleted.

The results of the factor analysis for the power, relationship commitment, customer integration and performance are shown in Figures 3 and 4. The measurement items all had strong loadings on the construct that they were supposed to measure and lower loadings on the constructs that they were not supposed to measure, indicating unidimensionality. The Cronbach's alpha values are shown in Figure 5. They were all above 0.80, except instrumental relationship commitment, which had an alpha value of 0.667, which is still above the lower limit of 0.60 suggested by Flynn et al. (1990) and Nunnally (1994) for newly developed scales. Although these scales were initially developed by Brown, et al. (1995), they had previously been applied only in Western countries. Thus, we applied the criterion for newly developed scales, finding that our measurements were unidimensional and reliable.

Construct Validity

We assessed construct validity along several dimensions. Convergent validity is the degree to which multiple methods of measuring a variable provide the same results, while discriminant validity is the degree to which measures of different latent variables are unique (O'Leary-Kelly & Vokurka, 1998). O'Leary-Kelly and Vokurka

Figure 3. Factor analysis of power

	Factor Loadings				
	Coercive Power	**Legitimate Power**	**Expert Power**	**Referent Power**	**Reward Power**
COE2	**.901**	.096	.012	.070	.125
COE4	**.870**	.106	.019	.059	.143
COE3	**.862**	.142	.043	.024	.227
COE1	**.827**	.087	-.044	.073	.204
LEG1	.076	**.777**	.238	.136	.112
LEG2	.118	**.773**	.115	.184	.145
LEG3	.107	**.725**	.215	.155	.158
LEG4	.117	**.697**	.118	.046	.367
EXP2	.019	.099	**.806**	.196	.169
EXP1	-.026	.177	**.787**	.104	.126
EXP3	-.102	.119	**.777**	.271	.102
EXP4	.160	.301	**.616**	.164	.088
REF2	.099	.162	.199	**.858**	.166
REF1	.062	.135	.248	**.826**	.153
REF3	.067	.204	.241	**.799**	.159
REW2	.212	.290	.161	.151	**.767**
REW3	.186	.261	.277	.158	**.738**
REW4	.347	.081	.161	.220	**.645**
REW1	.308	.439	.018	.159	**.589**
Eigenvalue	**3.389**	**2.843**	**2.675**	**2.420**	**2.357**
Total Variance Explained	72.018%				

(1998) suggest that confirmatory factor analysis (CFA) is a more powerful tool for assessing convergent and discriminant validity, requiring fewer assumptions than the traditional Multi-Trait Multi-Method matrix method. In our study, we tested both kinds of validity using CFA models.

In the convergent validity test, we constructed a CFA model using LISREL. Each item was linked to its corresponding construct, and the covariances among those constructs were freely estimated. The model fit indices were Chi-Square= 3786.98 with d.f.= 1130, RMSEA=0.065, NNFI=0.95, CFI=0.95 and standardized RMR=0.056, indicating that the model is acceptable (Hu & Bentler, 1992). A construct with either a loading of indicators of at least 0.5, a significant t-value, or both, is considered to be convergently valid (Fornell & Larcker, 1981; Chau, 1997). All factor loadings were greater than 0.50 and all t-values were greater than 2.0, therefore, convergent validity was demonstrated. In order to assess discriminant validity, we built a constrained CFA model in which the correlations among constructs were fixed to 1. This model was compared with the original unconstrained model, in which the correlations among constructs were freely estimated. A significant difference of Chi-square statistics between the fixed and unconstrained models indicates high discriminant validity (Fornell & Larcker, 1981; Chau, 1997). All the differences of χ^2 were significant at the 0.01 level, therefore, discriminant validity was demonstrated.

Results

We used structural equation modeling (SEM) to estimate the causal relationship among the constructs. SEM examines the relationships between

Figure 4. Factor analysis of integration, commitment and performance

	Factor Loadings				
	Customer Integration	**Normative Relationship Commitment**	**Customer Service**	**Financial Performance**	**Instrumental Relationship Commitment**
CI3	**.779**	.133	-.019	.062	-.024
CI10	**.760**	.061	.115	.070	.049
CI11	**.750**	.124	.113	.060	.079
CI8	**.734**	.089	.168	.005	.145
CI9	**.709**	.167	.221	.014	.166
CI2	**.663**	.070	-.069	.094	-.131
CI1	**.653**	.049	-.011	.048	-.148
CI5	**.621**	.132	.256	.111	-.094
CI4	**.619**	.172	.322	.117	-.049
CI7	**.607**	.125	.328	.119	-.011
CI6	**.585**	.065	.261	.129	-.073
NRC4	.149	**.846**	.136	.026	.146
NRC5	.126	**.840**	.143	.089	.134
NRC3	.140	**.800**	.176	.042	.106
NRC6	.174	**.754**	.162	.023	.196
NRC1	.124	**.699**	.256	.120	-.054
NRC2	.134	**.694**	.155	.082	-.064
CS4	.108	.156	**.794**	.021	.004
CS5	.077	.141	**.786**	.057	.050
CS6	.123	.179	**.732**	.074	-.027
CS1	.217	.209	**.706**	.189	.005
CS3	.278	.196	**.667**	.204	.076
CS2	.280	.183	**.558**	.177	.039
FPERF2	.076	.071	.107	**.872**	.019
FPERF4	.094	.079	.113	**.862**	-.042
FPERF5	.115	.068	.078	**.836**	.025
FPERF3	.133	.046	.133	**.802**	-.083
FPERF1	.095	.066	.119	**.797**	.050
IRC2	.044	.171	.016	-.023	**.820**
IRC1	-.124	-.005	-.052	-.018	**.788**
IRC3	-.005	.129	.086	.017	**.632**
Eigenvalue	**5.564**	**4.008**	**3.751**	**3.714**	**1.926**
Total Variance Explained	61.171%				

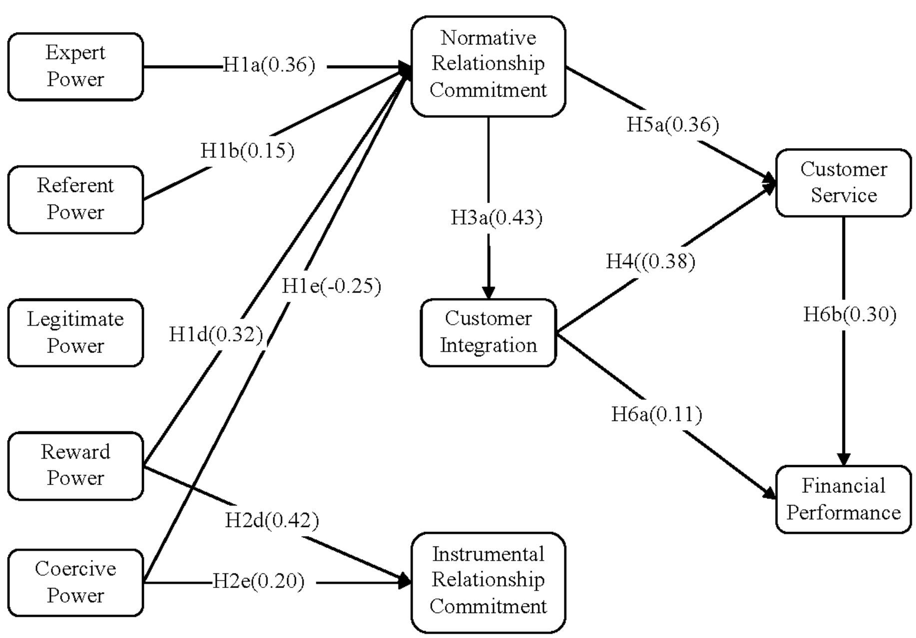
Figure 5.Reliablity analysis

Construct	**# of Items**	**Cronbach's Alpha**
Financial Performance	5	0.905
Normative Relationship Commitment	6	0.897
Instrumental Relationship Commitment	3	0.667
Expert Power	4	0.813
Referent Power	3	0.875
Legitimate Power	4	0.825
Reward Power	4	0.831
Coercive Power	4	0.915
Customer Integration	11	0.900
Customer Service	6	0.861

variables as a unit, rather than piecemeal, as in regression, and the assumption of perfectly reliable measures is untenable in regression, while it is easily handled by SEM (Kline and Klammer, 2001). SEM is a confirmatory approach to data analysis, requiring prior assignment of intervariable relationships. It statistically tests a hypothesized model to determine the extent to which the proposed model is consistent with the sample data. The measurement models specify how latent variables are measured in terms of indicator variables, as well as addressing the reliability and validity of the indicator variables in measuring the latent variables or hypothetical constructs. SEM provides an assessment of predictive validity, specifies direct and indirect relations among the latent variables, and describes the amount of explained and unexplained variance in the model (Byrne 1998; Schumacker & Lomax, 1996).

LISREL 8.54 was used to analyze the hypothesized structural equations model. A two-step model building approach was used, wherein the measurement models were tested prior to testing the structural model (Joreskog & Sorbom, 1993; Anderson & Gerbing, 1988). The Maximum Likelihood Estimation (MLE) method was used, because it has desirable asymptotic properties (e.g., minimum variance and unbiasedness) and is scale-free. This method assumes multivariate normality of the observed variables (Raykov & Marcoulides, 2000), and the requirement of normality of the data was met in our data. The structural model was built on the modified measurement model using the MLE method. In structural equation modeling, there is no single test of significance that can absolutely identify a correct model given the sample data (Schumacker & Lomax, 1996), and many goodness of fit criteria has been established (Bentler 1992; Garver & Mentzer, 1999, Cheung & Rensvold, 2001). The goodness of fit indices for our model are Chi-Square=3932.53 with d.f=1148, RMSEA=0.066, NNFI=0.94, CFI=0.95, Standardized RMR=0.068, which are better than the threshold values suggested by Hu and Bentler (1992). Therefore, our model can be accepted. Figure 6 shows the modified structural

Figure 6. Structural equation mode



equation model and the standardized coefficients for the paths that were significant at the 0.05 level.

From Figure 6, we can see that expert power and referent power were positively related to normative relationship commitment, indicating that these types of non-mediated power by the customer are related the manufacturer's commitment, due to identification with and internalization of customer's norms and values. This supports H1a, and H1b. However, the hypothesized relationship between legitimate power and normative relationship commitment was insignificant, and H3c was not supported. Expert power, referent power, and legitimate power were not related to instrumental relationship commitment, which does not support the hypothesized negative relationship between non-mediated sources of power and instrument relationship commitment; this is consistent with Brown, et al.'s (1995) finding that non-mediated power had no or a negative impact on instrumental relationship commitment. In China, when manufacturers perceive a customer to have expert and referent power, they tend to accept them without suspicion about the authority of the power sources. Therefore, non-mediated power was associated with normative relationship commitment, not with instrumental relationship commitment. In other words, the expert and referent power of customers was related to manufacturers' commitment in a normative way and restrains it from being exercised in an instrumental way. The strong relationship between expert power and normative relationship commitment indicates that Chinese people have strong belief in knowledge and authority. The relatively weaker relationship between referent power and normative relationship commitment shows that it plays a less important role in interfirm relationships in China. Legitimate power was not significantly related to either type of relationship commitment, perhaps because Chinese companies pay little attention to the influence of legitimate power in their high power distance culture.

The path coefficients in Figure 6 show that customer reward power significantly related to both normative and instrumental relationship commitment; thus H1d was not supported, but H2d was. Coercive power was positively related to instrumental relationship commitment, but negatively related to normative relationship commitment, thus supporting both H1e and H2e. This shows that punishment plays a significant role in improving instrumental relationship commitment, but is associated with decreasing normative relationship commitment. It is interesting that coercive power was negatively associated with normative relationship commitment, while reward power was positively associated with it, since both reward and coercive power are mediated sources of power. This may reflect the Chinese tendency to use positive feedback to encourage others to commit to their values and norms, while using negative feedback to regulate and manage calculative relationship commitment. The positive relationship between reward power and normative relationship commitment contradicts Brown, et al.'s (1995) finding that mediated sources of power were negatively associated with normative relationship commitment. This might be caused by cultural differences between China and the U.S. (Hofstede, 1983, 1984), particularly in terms of power distance, which is much higher in China than in the U.S. In high power distance cultures, mediated power is more likely to influence others. To confirm our interpretation of these findings, future cross-cultural studies should be carried out to examine the moderating effect of national culture on the relationship between power and relationship commitment.

The path coefficients in Figure 6 show that normative relationship commitment was strongly related to customer-manufacturer integration, supporting H3a. However, the coefficient for the path from instrumental relationship commitment to customer integration was not significant and thus does not support H3b. Comparing the equal coefficients constrained model with the unconstrained model, we find that the two coefficients were significantly different from each other, indicating support for H3c. These results are consistent with earlier finding that relationship commitment positively influences acquiescence and cooperation among channel members (Morgan & Hunt, 1994). However, our study indicates that there was a significant difference between normative and instrumental relationship commitment and their relationship with SCI. Since SCI requires investment of resources in sharing information and improving inter-organizational processes, partners must have a longer-term orientation, as well as congruence in their values, norms of behavior and managerial approaches. Therefore, normative commitment is required, to enhance customer integration. Figure 6 also indicates that the degree to which a manufacturer is integrated with its major customer was strongly related to manufacturers' customer service and financial performance, supporting H4 and H6a. The strong positive relationship between customer integration and performance indicates the importance of SCI in enhancing manufacturers' competitiveness. This finding is supported by previous studies that found that SCI could enhance performance (e.g., Frohlich & Westbrook, 2001; Stank, Keller & Closs, 2001).

The path coefficients for normative relationship commitment show that it strongly is positively related to manufacturers' customer service performance, supporting H5a. The coefficient for instrumental relationship commitment was insignificant, therefore it did not support H5b. Furthermore, through a comparison of the equal coefficients constrained and unconstrained models, we find that the impact of normative relationship commitment on customer service was significantly higher than the corresponding impact of instrumental relationship commitment. Therefore, H5c was supported. The result that normative relationship commitment was related to customer service was consistent with Brown, et al.'s (1995) finding that retailers' normative relationship commitment was positively associated with perceived performance of their suppliers. This is because normative relationship commitment improves congruence in values, management philosophies and norms of behavior. Furthermore, normative relationship commitment enhances acquiescence and cooperation. Therefore, when a manufacturer has greater normative relationship commitment to its major customer, it performs better in customer service, due to cooperative behaviors between them. The finding that instrumental relationship commitment was not related to customer service appears to be consistent with Brown, et al.'s (1995) finding that instrumental relationship commitment did not negatively influence the retailer's perceived performance of the supplier. Figure 6 also indicates that customer service had a positive impact on the financial performance of the manufacturer, supporting H6b. This finding highlights the importance of customer service performance in improving manufacturers' financial performance.

DISCUSSION

This study investigated the relationship between power, relationship commitment, customer integration and manufacturers' customer service and financial performances, based on data collected from manufacturers in China. A holistic model was built, allowing for simultaneous testing of these relationships by SEM analysis. China's manufacturing industries are growing rapidly, and our findings provide significant managerial implications for supply chain practitioners and researchers. In this section, we discuss our major findings and their managerial implications.

How to Improve the Performance of a Manufacturer in the Supply Chain?

Our model shows that the key drivers of manufacturers' financial performance are their customer service and the extent of customer integration. In this study, customer service was mainly measured by delivery and flexibility. Although delivery and flexibility have been cited as important competitive priorities in the operations management literature (Aggarwal, 1997), our study is the first to demonstrate their strong impact on manufacturers' financial performance in China. This indicates that manufacturers in China should emphasize flexibility and delivery, in order to achieve better financial results. Due to keen competition, cost and quality are increasingly becoming order qualifiers for many manufacturers in China. Companies need greater flexibility in meeting the changing needs of their customers and delivering products faster and more dependably, in order to win orders.

Our results also show that customer integration significantly enhanced manufacturers' customer service and financial performance. These results provide strong empirical support for the importance of SCI. Through information sharing and process integration with customers, manufacturers can be more flexible by introducing new products faster, modifying the existing product mix and adjusting production volumes faster, to better meet the needs of their major customers. Integration with customers also allows manufacturers to deliver products faster and more reliably. Therefore, manufacturers should invest time and resources to integrate with their major customers.

What is the Role of Relationship Commitment in Customer Integration and Performance Improvement?

Our study is the first to examine the impact of relationship commitment on SCI. The results show that manufacturers' normative relationship commitment was positively related to the degree to which manufacturers are integrated with customers, while instrumental relationship commitment was not related. Thus, manufacturers should cultivate normative relationship commitment with their customers, in order to enhance integration. With normative relationship commitment, customers cooperate with manufacturers and thus share information and integrate inter-organizational processes and activities. In contrast, due to the short term and loose nature of instrumental relationship commitment, it does not have any significant influence on customer integration.

Normative relationship commitment was also positively related to manufacturers' customer service, however, instrumental relationship commitment was not related to it. This shows that relationship commitment can result in improved customer service performance, due to cooperative behaviors and reduced conflict between partners. There was no significant effect of instrumental relationship commitment on customer service. Therefore, we see that normative relationship commitment is much more effective than instrumental relationship commitment in enhancing manufacturers' customer service. As we discussed in the previous section, customer service is significantly related to financial performance. As a result, normative relationship commitment has a higher impact on manufacturers' financial performance than instrumental relationship commitment. This indicates that normative relationship commitment plays a very important role in improving relationship commitment to customers, in order to achieve better financial performance.

How Does Power Influence Relationship Commitment?

Our model shows that the customer power is significantly related to relationship commitment. Customers' expert and referent power were positively related to manufacturers' normative relationship commitment, but not to their instrumental relationship commitment. Customers' use of legitimate power was not significantly related to either type of relationship commitment. Expert power was the most important, among the five types of power, in its relationship to normative relationship commitment. Therefore, customers should strive to enhance their expert power by hiring knowledgeable people and managing their expertise and skills to influence their suppliers. The use of legitimate power, however, is not encouraged, because there is no significant relationship between it and either type of relationship commitment.

When customers use coercive power to influence manufacturers, the manufacturer's instrumental relationship commitment is enhanced, while its normative relationship commitment is reduced. Given the strong, positive impact of normative relationship commitment on customer integration and manufacturers' performance, customers should refrain from the use of coercive power. It is interesting that customers' use of reward power had a positive significant impact on both normative and instrumental relationship commitment. On one hand, customers' use of reward power was associated with manufacturers' instrumental relationship commitment. On the other hand, however, reward power was also associated with manufacturers' normative relationship commitment. Thus, it may be that reward power is used for different purposes in China.

Our model shows that customers should emphasize the use of non-mediated sources of power, to enhance normative relationship commitment. With increased normative relationship commitment, manufacturers will be able to serve their customers better in terms of flexibility and delivery, which leads to better financial performance. In contrast, the use of mediated sources of power by the customer will be ineffective or will produce negative effects on the manufacturer's relationship commitment and, thus, will not improve performance. Partners within the supply chain must understand the effect of different types of power, and should selectively exercise their power, in order to enhance relationship commitment and improve performance.

CONCLUSION

Our study proposes and validates a model of the relationship between power, relationship commitment, customer integration and performance in a supply chain. It shows that the use of expert and referent power are significantly related to normative relationship commitment, while legitimate power was not related to it, and all three types of non-mediated power were not related to instrumental relationship commitment. Coercive power was related to instrumental relationship commitment but negatively associated with normative relationship commitment, while the use of reward power was positively associated with both normative and instrumental relationship commitment. Furthermore, the model shows that expert power had the strongest impact on normative relationship commitment. This is the first empirical study to systematically examine the relationship between power and relationship commitment in a supply chain. These findings provide a guideline for managers in understanding power in supply chain relationships.

Our model demonstrated that normative relationship commitment is strongly related to customer integration and manufacturers' customer service performance in a supply chain, clearly showing the importance of managing supply chain relationships. When partners have the intrinsic desire to continue a relationship, due to congruence in values and norms, SCI can be accomplished more readily, and the resulting commitment can help enhance performance. However, instrumental relationship commitment is much less effective in enhancing SCI and operational performance. Therefore, companies should strive to improve normative relationship commitment.

Finally our model shows that customer service, in terms of delivery, flexibility and level of customer service, has a strong impact on financial performance. Our study provides empirical support for an increased emphasis on flexibility and delivery as competitive priorities in today's global competitive environment. This indicates that Chinese manufacturers which focus on these priorities will achieve better financial performance. Being on the receiving end of non-mediated power from the partner, enhancing normative relationship commitment and increasing the level of integration are also effective ways of improving financial performance.

In addition to ample implications for practicing managers, this study also makes significant contributions to the supply chain management and relationship management literature. We demonstrated the importance of SCI and normative relationship commitment in improving customer service. We empirically showed that customer service is very important in improving manufacturers' financial performance. We demonstrated how the use of different types of power influences different types of commitment. We also found that some of the relationships between power and relationship commitment in China were different from those reported by Brown, et al. (1995), based on studies using data from the U.S. While Brown, et al. (1995) reported that mediated power was negatively related to normative relationship commitment, we found that reward power was positively related to both normative and instrumental relationship commitment in China. We speculate that these differences might be caused by the differences in national culture between China and the U.S. Further studies are needed to examine cross-cultural differences in the relationship between power and relationship commitment, and this study opens up a new stream of research in this area. Finally, the instrument validated by this study can be used in future studies and will thus foster future research in this important area.

Although the study described in this chapter makes significant contributions to both academia and practice, there are several limitations which open up venues for further research. First, this study did not examine the impact of power asymmetry on the relationships between power, relationship commitment, SCI and performance, which may moderate some of these relationships (Brown, et al., 1995), and thus should be included in future studies. Related to this, the impact of industry should also be investigated. In some industries, the balance of power between manufacturers and customers might be different and, therefore, the relationship between power, relationship commitment and performance will be different. The most important competitive priorities may vary by industry, due to differences in customer requirements and preferences. In addition, we only used data from China to develop and test this model. Because culture may have a significant influence on the role of power and relationship commitment in inter-firm relationships, this study should be extended to examine the moderating effect of culture on the relationships focused upon in this study.

While power and relationship commitment were found to be related to SCI, many other factors, such as competitive hostility and environmental uncertainty, may also be related to SCI and the impact of SCI on performance. Future studies should seek more drivers of SCI and examine their impact. One potentially important factor to be considered is trust and how different types of trust influence SCI. This study only examined the power of the major customer, based on perceptual data from manufacturers. Future studies should collect the perspectives of both manufacturers and customers. Comparison of both perspectives may shed more light on the relationship between power and relationship commitment. Finally, to understand the entire supply chain, future studies should examine power and relationship commitment among suppliers, manufacturers and customers together. Comparison of how they are related to SCI and performance among all three parties will reveal more complex dynamic relationships among these variables.

REFERENCES

Aggarwal, S. (1997). Flexibility management: the ultimate strategy. *Industrial Management (Des Plaines)*, *39*(1), 26–31.

Ahmad, S., & Schroeder, R. G. (2001). The impact of electronic data interchange on delivery performance. *Production and Operations Management*, *10*(1), 16–30. doi:10.1111/j.1937-5956.2001.tb00065.x

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. doi:10.1037/0033-2909.103.3.411

Armstead, C. G., & Mapes, J. (1993). The impact of supply chain integration on operating performance. *Logistics Information Management*, *6*(4), 9–14. doi:10.1108/09576059310045907

Beamon, B. M. (1999). Measuring supply chain performance. *International Journal of Operations & Production Management*, *19*(3), 275–292. doi:10.1108/01443579910249714

Bender, D. (1986). Financial impact of information processing. *Journal of Management Information Systems*, *3*(2), 22–32.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the bulletin. *Psychological Bulletin*, *112*(3), 400–404. doi:10.1037/0033-2909.112.3.400

Benton, W. C., & Maloni, M. (2005). The influence of power driven buyer/seller relationships on supply chain satisfaction. *Journal of Operations Management*, *23*, 1–22. doi:10.1016/j.jom.2004.09.002

Bowersox, D.J., Closs, D.J. & Stank, T.P. (1999). *21st Century Logistics: Making Supply Chain Integration a Reality*. Oak Brook, IL: Council of Logistics Management.

Bowersox, D. J., & Morash, E. A. (1989). The integration of marketing flows in channels of distribution. *European Journal of Marketing*, *23*(20), 58–67. doi:10.1108/EUM0000000000546

Boyer, K. K. (1999). Evolutionary patterns of flexible automation and performance: a longitudinal study. *Management Science*, *45*(6), 824–842. doi:10.1287/mnsc.45.6.824

Boyer, K. K., Leong, G., Ward, P., & Krajewski, L. (1997). Unlocking the potential of advanced manufacturing technologies. *Journal of Operations Management*, *15*(4), 331–347. doi:10.1016/S0272-6963(97)00009-0

Boyer, K. K., & Lewis, M. W. (2002). Competitive priorities: investigating the need for trade-offs in operations strategy. *Journal of Operations Management*, *11*(1), 9–20.

Brown, J. R., Lusch, R. F., & Muehling, D. D. (1983). Conflicts and power-dependence relations in retailer-supplier channels. *Journal of Retailing*, *59*(4), 53–80.

Brown, J. R., Lusch, R. F., & Nicholson, C. Y. (1995). Power and relationship commitment: Their impact on marketing channel member performance. *Journal of Retailing*, *71*(4), 363–392. doi:10.1016/0022-4359(95)90019-5

Bruggen, G. H., Van Lilien, G. L., & Kacker, M. (2002). Informants in organizational marketing research: Why use multiple informants and how to aggregate responses. *JMR, Journal of Marketing Research*, *39*(4), 469–478. doi:10.1509/jmkr.39.4.469.19117

Byrne, B. (1998). *Structural Equation Modeling With Lisrel, Prelis, and Simplis: Basic Concepts, Applications, and Programming*. Upper Saddle River, NJ: Lawrence Erlbaum Associates.

Chang, T. Z., & Chen, S. J. (1998). Market orientation, service quality and business profitability: a conceptual model and empirical evidence. *Journal of Services Marketing*, *12*(4), 246–264. doi:10.1108/08876049810226937

Chau, P. Y. K. (1997). Reexamining a model for evaluating information center success using a structural equation modeling approach. *Decision Sciences*, *28*(2), 309–334. doi:10.1111/j.1540-5915.1997.tb01313.x

Chen, I. J., & Paulraj, A. (2004). Understanding supply chain management: critical research and a theoretical framework. *International Journal of Production Research*, *42*(1), 131–163. doi:10.1080/00207540310001602865

Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods*, *4*, 235–263. doi:10.1177/109442810143004

Christiansen, T., Berry, W. L., Bruun, P., & Ward, P. (2003). A mapping of competitive priorities, manufacturing practices, and operational performance in groups of Danish manufacturing companies. *International Journal of Operations & Production Management*, *23*(10), 1163–1183. doi:10.1108/01443570310496616

Cox, A. (2001). Understanding buyer and supplier power: A framework for procurement and supply competence. *Journal of Supply Chain Management*, *37*(2), 8–15. doi:10.1111/j.1745-493X.2001.tb00094.x

Dangayach, G. S., & Deshmukh, S. G. (2001). Manufacturing strategy: Literature review and some issues. *International Journal of Operations & Production Management*, *21*(7), 884–932. doi:10.1108/01443570110393414

Daugherty, P. J., Ellinger, A. E., & Gustin, C. M. (1996). Integrated logistics: achieving logistics performance improvements. *Journal of Supply Chain Management*, *1*(3), 25–33. doi:10.1108/13598549610155297

Davis, M., Aquilano, N., & Chase, R. (2003). *Fundamentals of Operations Management* (pp. 37–39). New York: McGraw-Hill Irwin.

Dixon, J. R., Nanni, A. J., & Vollmann, T. E. (1990). *The New Performance Challenge- Measuring Operations For World-Class Competition*. Homewood, IL: Dow Jones-Irwin.

Eccles, R. G., & Pyburn, P. J. (1992). Creating a comprehensive system to measure performance. *Management Accounting*, *74*(4), 41–44.

Flynn, B. B., & Flynn, E. J. (2004). An exploratory study of the nature of cumulative capabilities. *Journal of Operations Management*, *22*(5), 439–458. doi:10.1016/j.jom.2004.03.002

Flynn, B. B., Sakakibara, S., Schroeder, R. G., Bates, K. A., & Flynn, E. J. (1990). Empirical research methods in operations management. *Journal of Operations Management*, *9*(2), 250–273. doi:10.1016/0272-6963(90)90098-X

Flynn, B. B., Schroeder, R. G., & Sakakibara, S. (1994). A framework for quality management research and an associated measurement instrument. *Journal of Operations Management*, *11*(4), 339–575. doi:10.1016/S0272-6963(97)90004-8

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research*, *18*(1), 29–50. doi:10.2307/3151312

Frazier, G. L., & Summers, J. O. (1986). Perceptions of inter-firm powers and its use within a franchise channel. *JMR, Journal of Marketing Research*, *23*(5), 169–176. doi:10.2307/3151664

French, R. P., & Raven, B. H. (1959). The Bases of Social Power. In Cartwright, D. (Ed.), *Studies in Social Power, 155- 164.Ann Arbor, MI: University of Michigan Press*. Ann Arbor, MI.

Frohlich, M. T., & Westbrook, R. (2001). Arcs of integration: An international study of supply chain strategies. *Journal of Operations Management*, *19*(2), 185–200. doi:10.1016/S0272-6963(00)00055-3

Garver, M., & Mentzer, J. T. (1999). Logistics research methods: employing structural equation modeling to test for construct validity. *Journal of Business Logistics*, *20*(1), 33–58.

Goodman, L. E., & Dion, P. A. (2001). The determinants of commitment in the distributor-manufacturer relationship. *Industrial Marketing Management*, *30*(3), 287–300. doi:10.1016/S0019-8501(99)00092-9

Hall, R. W. (1983). *Zero Inventories*. Homewood, IL: Dow-Jones Irwin.

Hammer, M. (1990). Reengineering work: don't automate, obliterate. *Harvard Business Review*, *4*, 104–112.

Hewett, K., & Bearden, W. O. (2001). Dependence, trust and relational behaviors on the part of foreign subsidiary marketing operations: implications for managing global marketing operations. *Journal of Marketing*, *65*, 51–66. doi:10.1509/jmkg.65.4.51.18380

Hochwarter, W. A., James, M., Johnson, D., & Ferris, F. R. (2004). The interactive effects of politics, perceptions and trait cynicism on work outcomes. *Journal of Leadership & Organizational Studies*, *10*(4), 44–57. doi:10.1177/107179190401000404

Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of International Business Studies (pre-1986), 14*, 75-89.

Hofstede, G. (1984). The cultural relativity of the quality of life concept. *The Academy of Management Review (pre-1986), 9*, 389-398.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351–362. doi:10.1037/0033-2909.112.2.351

Johnson, H. T., & Kaplan, R. S. (1987). *Relevance Lost - The Rise and Fall of Management Accounting*. Boston: Harvard Business School Press.

Johnson, J. L. (1999). Strategic Integration in Industrial Distribution Channels: Managing the Interfirm Relationship as a Strategic Asset. *Journal of the Academy of Marketing Science*, *27*(1), 4–18. doi:10.1177/0092070399271001

Joreskog, K., & Sorbom, D. (1993). *LISREL 8: Structural Equation Modelling With The Simplis Command Language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kasulis, J. J., & Spekman, R. E. (1980). A framework for the use of power. *European Journal of Marketing*, *14*(4), 180–191. doi:10.1108/EUM0000000004899

Kathuria, R. (2000). Competitive priorities and managerial performance: a taxonomy of small manufacturers, *Journal of Operations Management*, special issue on configurations in OM: taxonomies and typologies, *18*(6), 627-641.

Kline, T. J. B., & Klammer, J. D. (2001). Path model analyzed with ordinary least squares multiple regression versus Lisrel. *The Journal of Psychology*, *135*(2), 213–225. doi:10.1080/00223980109603692

Lambert, D. M., Robeson, J. F., & Stock, J. R. (1978). An appraisal of the integrated physical distribution management concept. *International Journal of Physical Distribution and Materials Management*, *9*(1), 74–88.

Lee, H. L., & Billington, C. (1992). Managing supply chain inventory: pitfalls and opportunities. *Sloan Management Review*, 65–73.

Loehlin, J. C. (1998). *Latent Variable Models* (3rd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

Maloni, M., & Benton, W. C. (2000). Power influences in the supply chain. *Journal of Business Logistics*, *21*(1), 49–73.

Min, S., & Mentzer, J. T. (2004). Developing and measuring supply chain management concepts. *Journal of Business Logistics*, *25*(1), 63–99.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, *58*, 20–38. doi:10.2307/1252308

Morris M.H. & Calantone R.J. (1991). Redefining the purchasing function: an entrepreneurial perspective, *International Journal of Purchasing and Materials Management*, Fall, 2-9.

Narasimhan, R., & Jayaram, J. (1998). Causal linkages in supply chain management: an exploratory study of north american manufacturing firms. *Decision Sciences*, *29*(3), 579–605. doi:10.1111/j.1540-5915.1998.tb01355.x

Narasimhan, R., & Kim, S. W. (2002). Effect of supply chain integration on the relationship between diversification and performance: evidence from Japanese and Korean firms. *Journal of Operations Management*, *20*(3), 303–323. doi:10.1016/S0272-6963(02)00008-6

Naylor, J. B., Naim, M. M., & Berry, D. (1999). Leagility: integrating the lean and agile manufacturing paradigms in the total supply chain. *International Journal of Production Economics*, *62*, 107–118. doi:10.1016/S0925-5273(98)00223-0

Nunnally, J. (1994). *Psychometric Methods*. New York: McGraw-Hill.

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*, 387–405. doi:10.1016/S0272-6963(98)00020-5

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method bias in behavioral research: A critical review of the. literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. doi:10.1037/0021-9010.88.5.879

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports of organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544. doi:10.1177/014920638601200408

Raykov, T., & Marcoulides, G. A. (2000). *A First Course in Structural Equation Modeling*. New Jersey: Lawrence Erlbaum and Associates.

Schumacker, R. E., & Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling*. New Jersey: Lawrence Erlbaum and Associates.

Skinner, W. (1971). The anachronistic factory. *Harvard Business Review*, *49*, 61–70.

Stank, T. P., Keller, S. B., & Closs, D. J. (2001). Performance benefits of supply chain integration. *Transportation Journal*, *41*(2), 31–46.

Stank, T. P., Keller, S. B., & Daughery, P. J. (2001). Supply chain collaboration and logistical service performance. *Journal of Business Logistics*, *22*(1), 29–48.

Stevens, G. C. (1989). Integrating the supply chain. *International Journal of Physical Distribution and Logistics Management*, *19*(8), 3–8. doi:10.1108/EUM0000000000329

Tan, K. C., Kannan, V. R., & Handfield, R. B. (1998). Supply chain management: supplier performance and firm performance. *Journal of Supply Chain Management*, *34*(3), 2–9.

Tedeschi, J. T., Schlenker, B. R., & Lindskold, S. (1972). *The Exercise of Power and Influence: The Source of Influence, The Social Influence Process* (pp. 287–375). Chicago: Aldine-Atherton.

Van Hoek, R. I. (1998). Measuring the unmeasurable - measuring and improving performance in the supply chain. *Journal of Supply Chain Management*, *3*(4), 187–192. doi:10.1108/13598549810244232

Vickery, S. K., Droge, C., & Markland, R. E. (1997). Dimensions of competitive strength in manufacturing: an analysis of competitive priorities in the furniture industry. *Journal of Operations Management*, *15*, 317–330. doi:10.1016/S0272-6963(97)00012-0

Vickery, S. K., Jayaram, J., Droge, C., & Calantine, R. (2003). The effects of an integrative supply chain strategy on customer service and financial performance: an analysis of direct versus indirect relationships. *Journal of Operations Management*, *21*, 523–539. doi:10.1016/j.jom.2003.02.002

Ward, P. T., Duray, R., Leong, G. K., & Sum, C. C. (1995). Business environment, operations strategy and performance: an empirical study of Singapore manufacturers. *Journal of Operations Management*, *13*, 99–115. doi:10.1016/0272-6963(95)00021-J

Wetzels, M., de Ruyter, K., & van Birgelen, M. (1998). Marketing service relationships: the role of commitment. *Journal of Business and Industrial Marketing*, *13*(4/5), 406–423. doi:10.1108/08858629810226708

White, G. P. (1996). A survey and taxonomy of strategy-related performance measures for manufacturing. *International Journal of Operations & Production Management*, *16*(3), 42–61. doi:10.1108/01443579610110486

Wisner, J. D., & Tan, K. C. (2000). Supply chain management and its impact on purchasing. *Journal of Supply Chain Management*, *36*(4), 33–42. doi:10.1111/j.1745-493X.2000.tb00084.x

ENDNOTE

1 Based on an article originally published as Zhao, X., Huo, B., Yeung, J.H.Y. and Flynn, B.B., "The Impact of Power and Relationship Commitment on Integration Between Manufacturers in a Supply Chain." *Journal of Operations Management*, 2008, *26*(3), 368-388.

APPENDIX CONSTRUCT MEASUREMENT

Financial Performance

FPERF1: Growth in sales.
FPERF2: Growth in profit.
FPERF3: Growth in market share.
FPERF4: Growth in ROI.
FPERF5: Growth in return on sales.

Customer Service

CS1: Our company can quickly modify products to meet our major customer's requirements.
CS2: Our company can quickly introduce new products into the markets.
CS3: Our company can quickly respond to changes in market demand.
CS4: Our company has an outstanding on-time delivery record to our major customer.
CS5: The lead time for fulfilling customer orders (the time which elapses between the receipt of customer's order and the delivery of the goods) is short.
CS6: Our company provides high level of customer service to our major customer.

Customer Integration

CI1: The level of linkage with major customer through information network.
CI2: The level of computerization for our major customer ordering.
CI3: The level of sharing of market information from our major customer.
CI4: The level of communication with our major customer.
CI5: The establishment of quick ordering system with our major customer.
CI6: Follow-up with our major customer for feedback.
CI7: The frequency of periodical contacts with our major customer.
CI8: Our major customer shares Point of Sales (PoS) information with us.
CI9: Our major customer shares demand forecast with us.
CI10: We share our available inventory with our major customer.
CI11: We share our production plan with our major customer.

Normative Relationship Commitment to Customer

NRC1: We feel that our major customer views us as being an important "team member", rather than our being just another supplier.
NRC2: We are proud to tell others that we are a supplier for this customer.
NRC3: Our attachment to this customer is primarily based on the similarity of our values and those of this customer.
NRC4: The reason we prefer this customer to others is because of what is stands for, its values.

NRC5: During the past year. Our company's values and those of the major customer have become more similar.
NRC6: What this customer stands for is important to our company.

Instrumental Relationship Commitment to Customer

IRC1: Unless we are rewarded for it in some way, we see no reason to expend extra effort on behalf of this customer.
IRC2: How hard we work for this major customer is directly linked to how much we are rewarded.
IRC3: Bargaining is necessary in order to obtain favorable terms of trade in dealing with this customer.

Customer's Use of Expert Power

EXP1: The people in the customer's organization knew what they are doing.
EXP2: We usually got good advice fro m our major customer.
EXP3: The customer had specially trained people who really knew what had to be done.
EXP4: Our major customer's business expertise made them likely to suggest the proper thing to do.

Customer's Use of Referent Power

REF1: We really admire the way our major customer runs their business, so we tried to follow their lead.
REF2: We generally wanted to operate our company very similar to the way we thought the major customer would.
REF3: Our company did what the customer wanted because we have very similar feelings about the way a business should be run.

Customer's Use of Legitimate Power

LEG1: It was our duty to do as the major customer requested.
LEG2: We had an obligation to do what the major customer wanted, even though it wasn't a part of the contract.
LEG3: Since they were the customer, we accepted their recommendations.
LEG4: The major customer had the right to expect us to go along with their request.

Customer's Use of Reward Power

REW1: IF we did not do what as the major customer asked, we would not have received very good treatment from them.
REW2: We felt that by going along with the major customer, we would have been favored on some other occasions.
REW3: By going along with the major customer's requests, we avoided some of the problems other suppliers face.
REW4: Our major customer often rewarded us to get our company to go along with their wishes.

Customer's Use of Coercive Power

COE1: The major customer's personnel would somehow get back at us if we did not do as they asked and they would have found out.

COE2: The major customer often hinted that they would take certain actions that would reduce our profits if we did not go along with their requests.

COE3: The major customer might have withdrawn certain needed services from us if we did not go along with them.

COE4: IF our company did not agree to their suggestions, the major customer could have made things more difficult for us.

This work was previously published in Managing Global Supply Chain Relationships: Operations, Strategies and Practices, edited by Barbara Flynn, Michiya Morita and Jose Machuca, pp. 191-218, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 69
Firm-Specific Factors and the Degree of Innovation Openness

Valentina Lazzarotti
Carlo Cattaneo University, Italy

Raffaella Manzini
Carlo Cattaneo University, Italy

Luisa Pellegrini
University of Pisa, Italy

ABSTRACT

This chapter investigates the topic of how open innovation is actually implemented by companies, according to a conceptual approach in which open and closed models of innovation represent the two extremes of a continuum of different openness degrees; though, these are not the only two possible models. By means of a survey conducted among Italian manufacturing companies, this chapter sheds light on the many different ways in which companies open their innovation processes. Four main models emerge from the empirical study, which are investigated in depth in order to understand the relationship between a set of firm-specific factors (such as size, R&D intensity, sector of activity, company organization) and the specific open innovation model adopted by a company.

INTRODUCTION

The concept of "Open Innovation" (OI) is often studied supposing an artificial dichotomy between closed and open approaches, whilst the idea of exploring different degrees and types of openness in a sort of continuum seems to provide a more interesting avenue (Chesbrough, 2003b). Prior research has highlighted that open innovation may be pursued in different ways, which are identifiable in terms of organisational form of acquisition or commercialization, number and typologies of partners, phases of the innovation process that are actually open, the direction of opennes (inbound and/or outbound) and governance (hierarchical or flat).

Moreover, previous research has also attempted to study the relationships among different OI models and several contextual factors, driven by the idea that these factors could explain or,

DOI: 10.4018/978-1-4666-1945-6.ch069

at least, characterize the companies' choices in terms of degree of openness. Lastly, different OI models, defined according to this concept (i.e. degree of openness and models within their specific context), have been analysed in some preliminary work in terms of their performance (Lichtenthaler, 2008; 2009).

The objective of this chapter is thus threefold: first, to provide evidence in support of Chesbrough's (2003b) theoretical proposition that businesses may be located along an Open Innovation Continuum, second, through the use of extensive study, to identify any potential intermediate states between the extreme points of the Continuum - Open and Closed Innovators - and, third, to identify the contextual factors that affect the choices firms make along the Open Innovation Continuum.

In particular, for the identification of the potential intermediate states in the OI Continuum, we focalized on two variables representing the openness degree, which are not still deeply investigated: (1) the number and type of partners (partners variety) and (2) the number and type of phases of the innovation process open to external contributions in and/or out (innovation phase variety). It should be noted that we assume that the innovation process is composed of different phases: idea generation (identification of a technology opportunity through scouting, monitoring, market analysis, trends analysis); experimentation (from the idea to the prototype); engineering (transforming the prototype into an industrial project); manufacturing (defining and organising the "plant"); commercialisation (planning of commercialisation and promotional activities).

The choices in terms of OI will be investigated in terms of those contextual factors whose role is still controversial (Lichtenthaler, 2008), or otherwise it can be better understood in light of the concept of openness suggested here. Our investigation was carried out in Italy, where empirical evidence about OI is still poor. However, there are many pressures, arising from institutions, too, towards the establishment of collaborative models (Global Business Summit, 2010). Thus, investigating if, how and with what results companies work together becomes a relevant issue for both Italian scholars and practitioners.

We would also like to emphasize that our endeavour to identify any in-between states along the Open Innovation Continuum is the first attempt to research this topic and that the subject indeed requires further research in order to better characterise such intermediate states.

The following sections are divided into subtopics: a description of the pertinent literature (so as to better understand the research questions we posed), a description of the empirical study we carried out and the methodology used, the main research results, a discussion of the results, conclusions and future research.

THEORETICAL BACKGROUND

The Theoretical Framework and the Research Questions

Traditionally, large firms relied on internal research and development (R&D) to innovate and, in many industries, large internal R&D labs were a strategic asset and firms could internally discover, develop and commercialize technologies. This approach has been labelled the "closed innovation model" (Chesbrough, 2003a). Although it worked well for quite some time, the current innovation landscape has changed. Due to labour mobility, increasing R&D costs, abundant venture capital and widely dispersed knowledge across multiple public and private organizations and the need for specialisation in knowledge production, enterprises can no longer afford to innovate on their own, but rather need to engage in alternative innovation practices. In this regard, Open Innovation (OI) represents an important innovation practice that can help firms to innovate without having to rely only on their in-house strengths. Since Chesbrough published his book in 2003, the concept of "Open Innovation" has received a considerable amount of attention from practitioners and researchers. A large number of studies are adopting this term to describe the phenomenon where firms rely increasingly on external sources of innovation, which means that ideas, resources and individuals flow in and out of organizations (Chesbrough, 2003a). While contributions are still growing (Gassmann, 2006; Enkel et al., 2009; Enkel et al., 2010), the debate on innovation management is enriched by studies that critically examine the Open Innovation concept by exposing its weakness and limitations (Dahlander & Gann, 2007; Trott and Hartmann, 2009). In particular, the concept of Open Innovation is criticized because of the widespread view that the concept highlights an artificial dichotomy between closed and open approaches. On the other hand, the idea of exploring different degrees and types of openness in a sort of continuum (i.e. the openness degree concept) seems to provide a more interesting and richer avenue to investigate, (Chesbrough, 2003b; Dahlander & Gann, 2007). Indeed, this view allows for a deeper and more real investigation into company behaviour and into the particular nature and context of innovation sources (Chesbrough, 2003b; Gassmann, 2006; Dahlander & Gann, 2007).

In any case, the era of open innovation has begun and many firms are opening their innovation process to the outside world (Enkel et al., 2009). The way the innovation process can be opened has been studied in management literature from a variety of perspectives. Although the perspective that has received most of the attention in the literature is undoubtedly the direction of openness, other approaches have also been investigated. More specifically, these look at the number and types of partners, the kind of governance in the innovation networks, and the organisational forms chosen to define the links among partners (high vs. low integration level).

As regards the perspective connected with the "direction of openness", three models of open innovation can be observed: the inbound, exploration or outside-in process, the outbound, exploitation or inside-out process, and the coupled process (Keupp & Gassman, 2009; Lichtentaler, 2008; Enkel, et al., 2009). Thanks to the outside-in process, firms aim at enriching the company's own knowledge base through the integration of external knowledge sourcing, and hence increase their innovativeness (Enkel et al., 2009). Through the inside-out process, firms aim at earning profits by bringing ideas to market, selling IP, and multiplying technology by transferring ideas to the outside environment, in order to bring ideas to market faster than they could through internal development (Enkel et al., 2009). The coupled process combines the two abovementioned processes to simultaneously gain external knowledge and bring ideas to market.

As regards the perspective connected with the "types of partners" (Enkel et al., 2009), literature has highlighted the interactive character of the innovation process, suggesting that innovators use ideas and knowledge of external actors in their innovation processes: firms rely heavily on their interaction with lead users, suppliers, and a range of institutions inside the innovation system (von Hippel, 1988; Lundvall, 1992; Brown & Eisenhardt, 1995; Szulanski, 1996). With each innovation source, an organization can achieve different intensity levels of collaboration (Laursen & Salter 2006; Keupp & Gassman 2009). Hence, it is possible to define different open innovation models depending on both breadth (i.e., the number of sources used for innovation activities) and depth (i.e., the intensity of collaboration with each source).

As regards the perspective connected with the "kind of governance" in the innovation networks, there are two dimensions which need to be considered (Pisano & Verganti, 2008): openness, i.e. a large number of involved partners and hierarchy, i.e. the level of 'democracy' in decision making. On the basis of two such aspects, four open innovation models emerge: (1) the open/

hierarchical model, in which anyone can offer ideas but only one company defines the problem and chooses the solution; (2) the open/flat model, in which anyone can generate ideas, and no one has the authority to decide what is or is not a valid innovation; (3) the closed/hierarchical model, in which a company selects certain participants and decides which ideas are to be developed; (4) the closed/flat model, in which a selected group offers ideas, while making critical decisions together.

As regards the "organisational forms" chosen to define the links among partners (high vs. low integration level), there are four technology sourcing modes that firms can adopt: corporate Venture Capital investments, non-equity alliances, equity alliances and acquisitions. Each form carries with it different implications in terms of the investing company's reversibility and commitment (Chiesa & Manzini, 1998; van de Vrande et al., 2006). More precisely, corporate Venture Capital investments and non-equity alliances are reversible to some extent and involve a relatively low level of commitment from the investing company, while equity alliances and acquisitions require a high level of commitment and are hardly reversible.

In our opinion, all these contributions share two aspects: on one hand, they have a common interpretation of open innovation, while, on the other, they have a weakness.

Regarding common interpretation, all these contributions share the understanding that the open innovation models which the firms follow are not exclusively open or closed, but rather show varying degrees of openness: i.e. between the two pure models – open or closed, which represent the two extremes of a continuum – there are many shades of grey (Chesbrough, 2003b). Indeed, according to Dahlaner and Gann (2007), the dichotomy between open vs. closed is artificial and it is necessary to explore different degrees and types of openness: this can yield more insight in understanding openness.

With regards to weakness, the perspectives used in the previous contributions are not exhaustive in explaining the open innovation models followed by firms. In other words, the latest literature still does not fully explain in what ways the degree of openness can happen. Indeed, to the best of our knowledge, the literature does not address the question whether some firms conduct open innovation in many phases of their innovation funnel and if others focus only on a very few of them; if this is the case, we must consider which phases of the innovation funnel are open or closed.

Hence, literature does not help companies to find the right balance between closed and open phases of the innovation funnel. Neither is it clear if the phases of the innovation funnel that are permeable are open to many or just a few partners. With few exceptions, it is not even clear if the involved partners are different in terms of typologies or not. For instance, De Backer et al. (2008) analyzed such a problem and found that universities and government research institutes are generally considered to be an important source of knowledge transfer for the innovation activities of companies, especially regarding more upstream/research activities.

On the basis of these premises, our objective is to contribute to the literature which sustains that business reality is not based on pure open innovation, but on companies that invest simultaneously in closed as well as in open innovation activities (Enkel et al., 2009) throughout the innovation funnel with different partners. Hence, we will introduce a new perspective that considers both the number/typology of partners and the number/typology of phases, in order to understand if such a perspective can confirm the existence of different models of open innovation. Within this context we will try to answer the following specific research questions:

- Do different firm-specific factors characterize the models of open innovation?
- Do such different models show a different level of innovative performance?

Figure 1 depicts the constructs of our theoretical framework. The operationalisation of each construct is reported in detail in the Appendix (all questions were measured on a four-point Likert scale to indicate the frequency of use, with 1 = disagreement and 4 = agreement).

As explained in the introduction, the main objective of the chapter is to provide empirical evidence to the notion of OI as a continuum, that is to say that different OI models may exist. Before characterizing them by means of our empirical analysis, in the following we will analyze what the literature says about the relationships between some contextual factors (i.e. R&D intensity, size, type of industry, approach to innovation, company's objectives for collaboration, managerial-organizational actions supporting open innovation) and OI models and their performance, by highlighting areas that are lacking which justify our subsequent empirical analysis. First, an analysis of the relationships between the firm-specific factors and the open innovation models will be made. Then, an analysis of the impact of the open innovation models on innovative performance will follow. Specifically, what is lacking in the literature has been highlighted for each of the relationships studied.

The Firm-Specific Factors and Open innovation

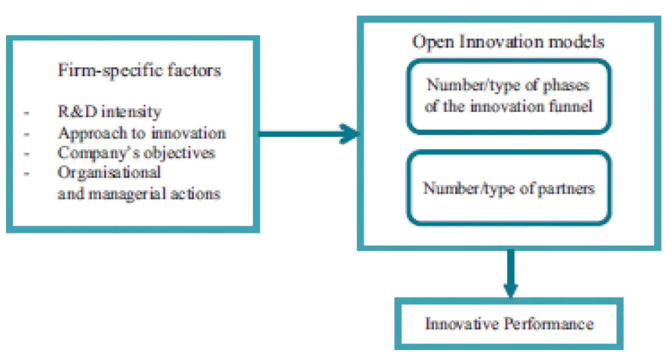
Relationship between R&D Intensity and Open Innovation Models

As regards the role played by R&D intensity, Lichtenthaler (2008), Lichtenthater and Ernst (2009), Calantone and Stanko (2007) and Sofka and Grimpe (2008) investigated this role from different viewpoints. Lichtenthaler (2008) and Lichtenthater and Ernst (2009) analysed the effect of R&D intensity and found that the greater the level of R&D intensity the greater the technological exploration. This provides support for the assumption that firms pursue external technology acquisition as a complement to internal R&D and not as a substitute (Cohen & Levinthal, 1990, Zahra & George, 2002).

Calantone and Stanko (2007) underpin that firms' exploration activities cannot occur frequently: therefore, given the high costs for developing specialized structures, firms are more likely to resort to outside expertise. Moreover, they state that firms performing a great deal of in-house exploratory research are likely to be led by this exploration away from their competencies, and will therefore be more likely to seek out outside expertise.

Sofka and Grimpe (2008) studied the effect of internal R&D investments on "breadth" (i.e. extent

Figure 1. Theoretical framework



of partner typology) and "depth" (i.e. intensity of collaboration for each partner) of research strategies. They hypothesize that internal R&D investments lead to deep research strategies rather than broad ones. With their survey, which involved firms from twelve European countries, they argue that firms building absorptive capacities through internal R&D have both broader and deeper research strategies. However, the effect on depth is stronger than the effect on breadth. In other words, committing internal resources to in-house labs and specialized scientists and engineers is therefore the primary path for innovation managers to achieve more depth in their search strategies.

Hence, on the basis of these contributions, the role played by R&D intensity is studied in relation to two of the abovementioned perspectives, through which it is possible to investigate how the innovation process can be opened: the perspective connected with the "direction of openness" and the perspective connected with the "types of partners". Thus, literature lacks the investigation of the role that R&D intensity plays with regards to the perspective offered in this chapter, which will consider both the variety of partners involved and the variety of stages in which companies collaborate.

Relationship between Size and Open Innovation Models

Size is one of the most investigated of the contextual factors and it is still a controversial subject.

On the one hand, empirical literature suggests that open innovation is mainly driven by larger companies. Empirical investigations show that size impacts on two variables representing the openness degree: the extent of both technology exploitation and exploration (Lichtenthaler, 2008; Lichtenthaler & Ernst, 2009). Indeed, as regards technology exploitation, larger companies seem to have a bigger technology portfolio than smaller companies and hence have wider technological knowledge that is potentially suitable for commercialization. As regards technology exploration, larger firms do not seem to be able to completely rely on internal activities due to the diversity of the technological knowledge that they use. In general, the fact that larger firms seem to drive the opening of the innovation process can be justified by the more systematic approach they have in their innovation processes (Lichtentaler, 2008) and the larger resources they possess with respect to small and medium enterprises (Lichtentaler, 2008; De Backer et al., 2008; van de Vrande et al., 2008). In addition, according to Lichtentaler (2008), it should be noted that the effect of size seems to be stronger in the case of technology commercialization than in technology exploration, in that commercialization is rather a newer phenomena than acquisition. As the external mode of technology exploitation has become a broader trend only in recent years, it is still driven by large pioneering firms, while the acquisition of external technology is distributed more evenly across firms of different sizes. Keupp and Gassman (2009) analyse the effect of size on two different variables representing the openness degree: the number of knowledge sources used for OI activities (i.e. the breath of OI) and the intensity of collaboration with each source (i.e. the depth of OI) and show that there is a positive and significant effect of firm size on both the breadth and the depth of OI.

On the other hand, some literature emphasises that especially small companies, often lacking resources and competence to innovate by themselves, would have great benefits from exploiting the OI model. In fact, SME are increasingly adopting OI practices (van deVrande et al., 2008).

Hence, on the basis of these contributions, it is possible to draw considerations similar to those regarding the role played by R&D intensity. Indeed, the role of size is studied in relation to the same two abovementioned perspectives: the perspective connected with the direction of openness and the perspective connected with the types of partners. In addition, its role is still controversial. As a consequence, it is possible to assert that literature lacks the investigation of the role that size plays with regard to the perspective offered in this chapter (number/typology of partners and number/typology of phases).

Relationship between Type of Industry and Open Innovation Models

Some authors investigate the impact that industry exerts on OI, interpreting industry as the typology of sector in which firms operate. Also in this case, literature is not unidirectional in that empirical findings show contrasting results. What's more, the same Lichtenthaler in different publications (Lichtenthaler, 2008 and Lichtenthaler & Ernst, 2009), while referring to the same sample, finds different results.

In fact, on the one hand, Lichtenthaler and Ernst (2009) show that a firm belonging to a particular industry does not produce any impact either on the external technology acquisition or on external technology exploitation. Similarly, Lichtenthaler (2008) states that his findings demonstrate the insignificant effect which industry differences have across the clusters. Thus, the openness of the innovation process does not seem to be determined principally by industry characteristics.

On the other hand, the studies by Gassman and Enkel (2004) and De Backer et al. (2008) demonstrate the opposite. More in particular, Gassman and Enkel (2004) state that the relative importance of internal and external sources varies across different industries. De Backer et al. (2008), although focusing on particular aspects, such as patent licensing, find important differences among industries, with chemical/drugs, electronic/electrical/semiconductors and machinery/equipment/computers as the industries where licensing deals take place more frequently than in others.

Relationship between Approach to Innovation and OI Models

A relevant concept investigated in the literature is that of "technology aggressiveness" (measured by three items, among them "the emphasis on radical innovation rather than incremental innovation")[1]. Lichtenthaler and Ernst (2009) find that technology aggressiveness is negatively related to the extent of external technology acquisition and is positively related to external technology exploitation, in that commercialization nurtures benefits in terms of setting industry standards, entering into new markets, and realizing learning effects.

In another publication, Lichtenthaler (2008) studies the implications connected with firms' emphasis on radical innovation and finds that the degree of openness seems to rise with the degree of emphasis on radical innovation, especially concerning the degree of external technology commercialization. There are two reasons for this: first, the opportunity to commercialize knowledge which, when not applied in the organization, turns out to be residual; second, the possibility to facilitate acceptance on the market and the creation of a standard. Lichtenthaler (2008) also finds that firms which emphasize radical innovation are obviously not able to develop all knowledge internally, but they have to strongly rely on complementary external sources and thus they use technology acquisition (Perrons et al., 2005).

Hence we can draw even more restrictive considerations than those regarding R&D intensity: technological aggressiveness is studied in the literature in relation only with the perspective connected with the direction of openness. If we add that this factor's role is still controversial, it emerges that new empirical investigation is needed to analyze the impact exerted by technological aggressiveness on OI models.

Company Collaboration Objectives and OI Models

The main reasons that push firms towards choices of open innovation are, on the one hand, the need to reduce innovation costs and business risks, and on the other, the need to extend skills, competences and creativity (Huang et al., 2009).

As regards the objective of diminishing costs and risks, Calantone and Stanko (2007) analyze outsourcing as a tool for increasing staffing efficiency measured in terms of employee sales efficiency. They infer that the decision to reduce the number of employees is related to the outsourcing of innovation in the short run but not over the long term. Gassmann and Enkel (2004) state that research-driven companies usually aim at reducing the R&D's fixed costs and sharing risk. Chiaroni, Chiesa and Frattini (2009) state that the reason for accessing external sources is the willingness to minimize risk by investing in technologies that are already proven in other applications.

Another main reason for firms to undertake R&D outsourcing includes accessing specialized skill sets and creativity, which exposes the internal development staff to new knowledge, technology, and organizational development processes (Catalone & Stanko, 2007; Chesbrough and Teece, 1996; Linder, 2004; Lynch, 2004), even if this strategy has drawbacks in terms of opening the market to new entrants (Porter, 1980) and exposing core competencies to imitation and substitution (Piachaud, 2005).

In comparison with the other firm-specific variables, the objectives of collaboration are studied in the literature even more restrictively; not only are they studied in relation to the perspective connected with the direction of openness, but also mainly in relation to one of the two directions, i.e. with the inbound process. Hence, for this firm-specific variable, too, there is a gap in the research literature which needs to be filled, that being an analysis of the impact exerted by collaboration objectives on the open innovation models.

Managerial-Organizational Actions Supporting Open Innovation and OI Models

Managerial-organisational actions allow open innovation to be pursued easily and more deeply, Some of these actions include the commitment of top management to promote the transition towards an open innovation approach (Vanhaverbeke, 2006; vandeMeer, 2007; Chiaroni et al., 2009, Pisano & Verganti, 2008); the need for a champion supporting the integration of external technology into an existing product development phase–gate process (Chesbrough, 2006; Chesbrough & Crowther, 2006); the exploitation of the personal relationship of the R&D managers for starting technological collaborations; the formal evaluation of collaboration objectives and risks, as well as the analysis and selection of the potential partners with a formal and explicit process (Sakkab, 2002; Huston & Sakkab, 2006). Although the works cited have shed light on how organizational and managerial factors facilitate the implementation of open models, we believe that enriching this line of inquiry with new empirical evidence is in any case quite important.

The Impact on Performance

The debate is still open on whether and how openness degree and contextual factors impact on innovative and economic performance. The results are still quite limited and contradictory, although very recent contributions (Chiang & Hung, 2010; Sofka & Grimpe, 2010) shed more light on the topic.

A widely accepted assumption is that the relations between openness degree and performance must be analyzed considering the moderator role of external environmental moderators (e.g. patent protection status: Lichtentaler, 2009; Slowinsky & Zerby, 2008; MacCormack & Iansiti, 2009). Indeed, regarding performance, it should be noted that the analysis of the company's financial performance is a complex topic due to the fact it can be explained only by considering a wide set of factors that can have contrasting effects.

Probably, the concept of innovative performance (Chiang & Hung, 2010; Sofka & Grimpe, 2010) is more understandable. The impact of open innovation models on innovative performance has been analysed in terms of a company's competence base, development costs, time to market and the level of innovation of new products/processes. Literature is unidirectional in showing the impact of the outside-in process on the access and integration of internal company capabilities with new and complementary knowledge of external firms (Gassmann and Enkel, 2004). Instead, literature results are not unidirectional as far as the reduction of development time is considered: for instance, on one hand, Gassmann and Enkel (2004) state that the benefits of co-operation are seen in an improvement in the competitive position and in risk minimisation, but not in a reduction of development time; on the other hand, according to Kolk and Püümann (2008) firms not concentrating on Open Innovation strategies fail, as rising development costs and shorter product life cycles make it increasingly difficult to justify investments in innovation. Other studies (e.g. Dahlaner & Gann, 2007) show that relationships with other actors help firms to increase the level of innovativeness.

In summary, as suggested by the literature listed above, certain relationships between the selected firm-specific variables and the openness degree are still controversial or lacking in depth. Below, we suggest improving the empirical evidence available by adopting a perspective based on number/typology of partners and the number/typology of phases, with particular reference to Italy, where partnerships are desired by many subjects, including institutional ones, though the issue is still poorly studied.

RESEARCH DESIGN AND METHODOLOGY

Survey Design

The empirical study has focused on companies located in Lombardy, a Northern Italian region; in 2008 the companies had applied for funding from the Chamber of Commerce to conduct innovative activities within different manufacturing sectors, including the mechanical and machinery sectors, as well as in sectors dealing with automotive, metallurgy, textiles, food, electronics, chemicals, pharmaceuticals, plastic, rubber, paper and paperboard, publishing and printing, wood and wood products (NACE rev.2 codes). This engagement in innovation by such companies, combined with the fact that Lombardy is marked by a particular propensity for innovation (if measured by the number of patents, Lombardy ranks first among the Italian regions according to the European Patent Office data for Italy elaborated by the Unioncamere Observatory of Patents and Brands, 2008) make them very interesting topics of innovation study. The data was collected by means of questionnaires distributed by email to participants. The advantages of such a method include low cost, completion at the respondent's convenience, absence of time constraints, guarantee of anonymity and reduction of interviewer bias (Forza, 2002). Its shortcomings, on the other hand, are represented by lower response rate as compared to other methods, longer completion times and greater effects due to the lack of both interviewer involvement and open-ended questions. The survey tool was conducted as a questionnaire whose items regarded company characteristics (sector, size), innovation strategy; organization for innovation; collaborations and innovative performance, as will be clarified in more detail below.

Before sending the questionnaire to the companies, a pilot test was conducted to assess the quality of the measure items. The items were tested by a group of senior managers and academics with working experience in innovation. They were asked to analyze the questionnaire in order to eliminate items not having strong content validity. After this stage, the resulting questionnaire was sent to the key informants of the companies that we identified as the R&D manager (if present in the company) or the company owner, if deeply involved in the definition of the company's innovation strategy (as is very common in Italian companies).

Statistical Analysis

Among the companies that have applied for funding (about 500) 99 firms have responded during a four-months period in 2009 (i.e. with a response rate equal to 20%). A general premise should be made as concerns company size (in terms of number of employees and revenues): except for few big subjects, the size of the studied companies can be classified as middle/small[2]. This imbalance can hardly be avoided because it is due to the intrinsic major sectoral composition in Lombardia, where the small size plays an important role. If, from the one hand, it is also found in non-respondents and thus it protects against the potential non-response bias, on the other hand it may prevent capture size differences when we will analyze firms characterized by different openness degrees. We must therefore bear in mind that this cannot make next comparisons between companies significant because of the intrinsic nature of the sample. However, this is the typical situation in Lombardia as well as in Italy. As clarified above in the theoretical background, we adopt the partner variety and the phase variety as relevant variables to represent and to investigate the degree of openness. Regarding their operationalization, we used subjective measures based on four-point Likert-type scales (1=strongly disagree; 4 = strongly agree) as given in the Appendix. In order to better specify the partner variety and the phase variety we introduced the variables: intensity of collaboration with partner and intensity of collaboration on phases that make up the innovation funnel (measures also based on the Likert-type scale). The study of correlations between these two variables allowed us to understand which are the most typical combination partner/phase and thus to characterize the innovation process in practice. To study firms' approaches to open innovation, we firstly carried out a cluster analysis ("complete linkage method", recommended when researcher wants to identify groups which are distinct from each other as much as possible; Barbaranelli, 2006) based on the partner variety and the phase variety. Secondly, concerning the firm-specific variables, with which we intended to describe the companies belonging to different clusters, we carried out the following procedure. Items of the questionnaire were defined on the basis of scales already used in previous works or coming from partial reworking of such scales (still Likert-type). Anyway, we applied to the gathered data an exploratory factor analysis (principal axis factoring as extraction method and promax rotation in the case of initially unclear solution) in order to delete weakly related items and to understand the factor structure and the measurement quality. An evaluation of the Eigenvalues and the Scree plot were used to identify the number of factors to retain. In addition, all factor loadings were above the acceptance level of 0.50 (Hair et al., 2006; Barbaranelli, 2006; Cheng and Shiu, 2008), thus indicating the unidimensionality of the various factors. These were saved as variables and employed in the subsequent analysis. The factors/firm-specific variables were the following (see the Appendix for detail):

Objectives of collaborations classified in two factors:

1. aims to extend skills, competences and creativity (three items, inspired by the work of Huang et al., 2009; Cronbach's α: 0,71);

2. aims to share risks and costs (two items, based on Calantone and Stanko, 2007; Cronbach's α: 0,84).

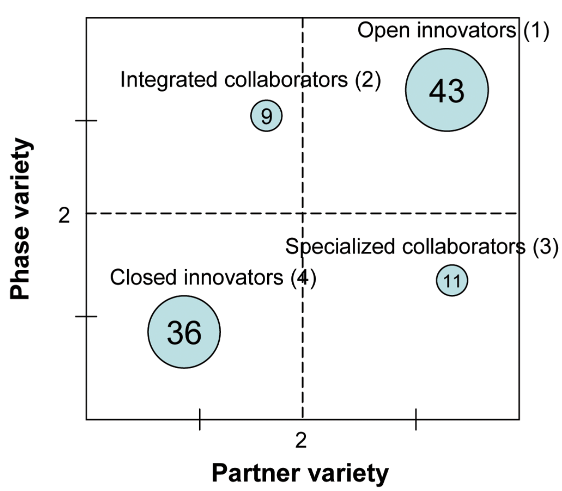
Approach to innovation: technological aggressiveness with emphasis on radical innovation (five items, inspired by Lichtenthaler and Ernst, 2009, that use suggestion by Brockoff and Pearson, 1992. We re-adapted the scale also on SURVEY TOOL 2.1 basis, a questionnaire sponsored by Industrial Research Institute, Cronbach's α: 0,71)

Organizational and managerial actions for open innovation (five items, scale based on SURVEY TOOL 2.1, Cronbach's α: 0,85)

Some other variables, not presented in Appendix, were measured directly (and eventually transformed in logarithmic scale to improve normality), such as:

- R&D intensity (i.e. percentage of R&D expenses/sales)
- Revenues (i.e. to operazionalize size)
- Number of employees (i.e. to operazionalize size)
- Indicators of company's performance (ROA – Return On Assets - and ROS – Return On Sales).

Figure 2. Results of cluster analysis



As concerns company's results, also a factor representing innovative performance was defined (five items reported in the Appendix, our scale based on Calantone et al. (2002); Cronbach's α: 0,82). After all, dummy variables were included for: the type of industry; the existence of organizational unit specifically devoted to support collaboration; the type of organizational structure used by companies for innovation activities (input-oriented, output-oriented, matrix; Chiesa, 2001).

Finally, as concerns data analysis, we applied the one-way variance analysis (i.e. ANOVA), in order to appreciate differences among clusters in terms of scale variables, and Chi-square test to compare the frequency on nominal variables.

RESULTS

Figure 2 illustrates the results of cluster analysis based on the partner variety and the phase variety.

This has resulted in a solution with four groups of firms. The decision on the number of clusters has been determined by the criterion that suggests of stopping the aggregation process at the stage that precedes the one with the highest increase in the coefficient of agglomeration (Barbaranelli, 2006). In the four-cluster solution, the variance inside clusters is about 21% whereas the variance among clusters is about 72% (F-tests sig. <.001). Despite the quite high correlation between partner variety and phase variety variables, which suggests that most firms adopt open or closed innovation approach on both dimensions, small intermediate clusters - 2 and 3 - exist (i.e. they open their innovation process more strongly in one direction rather than in the other) and thus they are worth to be analyzed although they only provide clues for further analysis.

Cluster 1 refers to the *open innovators*, companies that make up the largest group. From data about partner and phase variety variables (reported in Table 1), we found that these companies are really able to manage a wide set of technological relationships, that impact on the whole innovation funnel and involve a wide set of different partners. Although the open innovators strongly collaborate especially with the supplier in the engineering and experimentation phases[3], many other types of partners (particularly, firms operating in different sectors of activity, customers, universities, technical and scientific service companies, governmental institutions) are involved at different stages (especially in the idea generation, experimentation and engineering). An example is a medium-sized company from the machinery industry that conceptualizes and produces boilers for industrial use. Cluster 4 refers to the *closed innovators*, companies that form the second largest group. From data in Table 1, we realized that these companies access to external sources of knowledge only for a specific, single phase of the innovation funnel and typically in dyadic collaborations. The prevalent partners are in fact suppliers and customers, especially on the idea generation phase. Particularly little-used in the closed companies are the collaborations with universities and firms operating in different sectors of activity. A good example here is a small-sized textile company that has declared in some follow-up interviews, carried out after data-analysis stage, to follow a traditional innovation approach (i.e. internal research and development procedures "jealously preserved") by using the low-intensity contribution of its customers and suppliers on the idea generation phase. The companies in the smallest cluster 2 can be named *integrated collaborators*. Already found in the evidence emerged by a multiple case-study in a previous work (Lazzarotti & Manzini, 2009), these companies are the most similar to the closed ones: collaborations are with few types of partners (typically suppliers and customers), but instead of being tightly focused on one stage, they can be extended to the whole funnel. This means that the integrated collaborators "share" with their few trusted partners the whole process of innovation. An example is a small company in the electronics industry that produces and commercializes panels and electrical equipment: suppliers and customers, with whom the company has a longstanding relationship, support it through the whole process of innovation. Finally, companies in the other small cluster 4 can be classified as *specialized collaborators*. Already also emerged in our previous work, they form a group similar to open innovators regarding the variety of partners (suppliers, customers, universities, governmental institutions) but they concentrate their collaborations in a single/few points of the innovation funnel (typically the idea generation and the experimentation). From the follow-up interviews, these companies still seem a "bit behind the open", but it is a matter of time: with increasing confidence in partners, the cooperation will also increase by covering all the innovation process. An example here is a medium-sized company in the electronics industry that open to many partners (universities and governmental institutions as well as the traditional customers and suppliers) the idea generation phase.

To analyze differences across clusters, in particular the *open* and *closed* ones, one-way variance analysis (i.e. ANOVA) has been used for comparing means of scale variables (i.e. company's size expressed by revenues and number of employees, R&D intensity, approach to innovation, types of objectives, organizational and managerial actions for open innovation, performance expressed by innovative performance) and Chi-square test has been employed to compare the frequency on nominal variables (i.e. type of industry, existence of organizational units supporting collaborations). It must be said that the scarcity of observations on clusters *integrated* and *specialized* makes not applicable Chi-square tests as well as the results regarding scale variables are often not significant. Anyway, although the following evidence is useful above all to compare *open* and *closed* companies, some clues on other two clusters also emerge and thus, if interesting, they will be briefly presented in order to deepen them with next research.

Table 1. Information on the clusters and main differences (scale variables)

Variables	Sample (n=99)	Cluster 1 (n=43) Open	Cluster 2 (n=9) Integrated collaborator	Cluster 3 (n=11) Specialized collaborator	Cluster 4 (n=36) Closed	Significance (Anova test)
Partner and phase variables						
Partner variety	2.63	3.44	1.89	3.18	1.67	**.000**
Phase variety	2.61	3.49	3.11	1.82	1.67	**.000**
Intensity of collaboration with University and Research centres	1.38	1.59	1.16	1.36	1.18	**.005**
Intensity of collaboration with Technical and Scientific Service Companies	1.37	1.56	1.29	1.05	1.25	**.02**
Intensity of collaboration with Governmental institutions	1.11	1.20	1.00	1.18	1.01	**.03**
Intensity of collaboration with customers	1.70	1.77	1.56	1.80	1.61	.63
Intensity of collaboration with suppliers	1.93	2.12	1.80	1.87	1.74	.17
Intensity of collaboration with competitors	1.09	1.08	1.18	1.29	1.03	**.01**
Intensity of collaboration with firms operating in different sectors of activity	1.40	1.61	1.29	1.33	1.19	**.04**
Intensity of collaboration on Idea generation	1.51	1.62	1.38	1.52	1.39	.11
Intensity of collaboration on Experimentation	1.56	1.72	1.35	1.69	1.36	**.000**
Intensity of collaboration on Engineering	1.44	1.61	1.40	.1.32	1.29	**.002**
Intensity of collaboration on Manufacturing set up	1.34	1.49	1.13	1.26	1.22	**.002**
Intensity of collaboration on Commercialization	1.28	1.35	1.35	1.25	1.18	.23
Firm-specific contextual variables/factors						
Revenues (Log)	7.35	7.5	7.5	7.35	6.9	0.53
Employees (Log)	1.52	1.63	1.58	1.66	1.32	0.41
R&D intensity (Log)	0.59	0.86	0.20	0.59	0.35	**.01**
Innovation approach	-0.04	0.33	-0.07	0.30	-0.47	**.000**
Objective of extending skill and competence	-0.20	0.45	-0.13	-0.11	-0.47	**.000**
Objective of sharing risks and costs	0.02	0.28	-0.29	-0.13	-0.22	**.05**
Organizational and managerial actions for OI	0.01	0.37	0.26	0.34	-0.61	**.000**

continued on following page

Table 1. Continued

Variables	Sample (n=99)	Cluster 1 (n=43) Open	Cluster 2 (n=9) Integrated collaborator	Cluster 3 (n=11) Specialized collaborator	Cluster 4 (n=36) Closed	Significance (Anova test)
Performance						
Innovative performance	-0.04	0.30	0.08	0.29	-0.48	**.001**
ROS (Log)	0.73	0.76	0.80	0.65	0.70	0.89
ROA (Log)	0.78	0.78	0.69	0.75	0.81	0.94

Regarding *R&D intensity*, we found significant difference among open and closed clusters (see Table 1): the open companies invest more on average than the closed companies. This finding provides support for the assumption that firms consider open innovation as complementary with internal R&D and not a substitute (Lichtenthaler, 2008). This is consistent with the theory of absorptive capacity (Cohen & Levinthal, 1990, Zahra & George, 2002) in the sense that to be able to absorb from the outside, a company must have the appropriate skills and competences. This does not mean that the closed invests little, just it seems they invest less than open. As clues relatively to integrated and specialized clusters, interestingly the integrated shows the lowest average of R&D intensity whereas the specialist is more similar to the open one. Perhaps the integrated innovator invests internally less because it relies on a few trusted collaborators along the whole innovation process. In addition, in similar vein with previous absorptive capacity interpretation, Chi-square tests on internal organizational structure for innovation activities show that the input-oriented one (i.e. where people are organized according to their specific area of expertise, whose growth is continuously fed - Chiesa 2001) is typical for open innovators rather than for closed (for which an informal type of organization is prevalent). Indeed, the competence-building receives also a formally structured attention and thus it could suggest competence is considered a pre-requisite to openness.

As concerns *size* (see Table 1), we did not find significant differences among clusters. However, we reiterate that this may be due to the inherent imbalance of the investigated sample (i.e. high weight of medium-small sizes for companies) that reflects the Lombardia's and Italy's condition. But what perhaps we can say is that Italian companies, despite being small, are nevertheless brought to open up to outside sources, in keeping the stream of literature that argues that small firms behave like this (van de Vrande et al., 2008).

Groups do not seem different even for *type of industry* (see Table 2). As suggested by the follow-up interviews, the degree of open innovation seems to be mainly determined by the individual strategic choice of a company rather than by industry characteristics (for similar evidence, see Lichtenthaler, 2008).

As concerns *approach to innovation*, with emphasis on radical innovation rather than incremental innovation, we found that the open cluster has a higher mean (in the factor score resulting by factor analysis) than the closed and this difference is statistically significant. This is consistent with the literature and empirical evidence that suggest that companies, when focalized on radical innovations, must collaborate because they are not able to internally develop all relevant knowledge (Lichtenthaler, 2008; Perrons et al., 2005). "Relevant knowledge" that our data seem to suggest is coming from a higher degree of openness in term of wide partner variety and wide phase variety. Interestingly, the specialized, al-

Table 2. Information on open and closed and main differences (dummy variables)

Variables	Sample (n=99)	Cluster 1 (n=43) Open	Cluster 4 (n=36) Closed	Significance (Chi-square test)
Industry				
Mechanic/machinery	41%	46%	34%	0.29
Metallurgy	14%	12%	18%	0.38
Textile	8%	7%	9%	0.7
Food	4%	5%	3%	0.7
Electronics	7%	7%	6%	0.9
Chemical/pharmaceuticals	10%	14%	6%	0.42
Organizational context				
Organization input-oriented	44%	53%	33%	**.05**
Existence of organizational unit supporting collaboration	35%	47%	22%	**.02**

though very little, shows a mean higher than integrated, which might suggest that perhaps the partner variety is more relevant.

Regarding the *type of objectives* pushing companies to collaborate, open cluster shows higher mean (and statistical different) in the first-type goal "aim to extend skills, competences, creativity" with respect to closed companies. Very similar to each other (and in intermediate position between open and closed), the integrated and the specialized cluster. This finding suggests that companies look for competences and creativity by opening up in some way: to a wide variety of partners (even if on few phases), to a wide variety of phases (even if with few partners) or, at the highest level, in both directions. Similarly, the second-type goal of sharing risks and costs is related to the degree of openness in our conception. It is also interesting to note the prevalent objective in each cluster. Whereas in open cluster the main goal is the first, closed companies are pushed to open by the objective of sharing costs and risks.

As concerns the *organizational and managerial actions*, open cluster shows an average intensity on these tools statistically higher than closed (integrated and specialised still similar each other and in intermediate position between open and closed). As suggested by literature (Pisano & Verganti, 2008) these type of actions are necessary to ensure successful collaborations and this is confirmed by our evidence. Integrated and specialised have got a lower degree of openness (the integrated lower than the specialized) and probably a lower complexity in the collaborations. Thus, it is not necessary to introduce high-intensive managerial and organizational actions. Also a significant Chi-square test on the existence of an organizational unit supporting collaboration (see Table 2) gives evidence of the organizational and managerial differences between open and closed approaches.

Regarding *performance* (see Table 1) we obtained only some preliminary indications and mainly focus on innovative performance. Indeed, we believe that the analysis of overall company's performance is a complex topic due to the fact it can be "explained" only by considering a wide set of factors that can have opposite impacts. With this premise, we studied the differences between clusters only with an explorative purpose to define next steps of research and in terms of innovative performance (i.e. factor that is a combination of five items). We found that open cluster seems more performing than closed (and better than the sample average). Moreover, by studying correlations between innovative performance and partner variety, we found a high and significant relation. Particularly strong were the relations between the single item "The company's competence base was enlarged" and partner variety and "the level of innovativeness of new products/processes was improved" and still the partner variety, suggesting that the open is more innovative and that the innovativeness seems to be linked to the partner variety. Another clue for this type of interpretation is given by the specialist innovative performance: higher than integrated just in these two items. Anyway, it is important to keep in mind that they are only clues, not confirmed by the analysis of company's overall performance (measured by means of ROS and ROA), that is even greater in the closed than in the open cluster. Thus, further investigation on performance is surely required.

DISCUSSION

In this chapter, different models for open innovation are studied, by means of a survey conducted among 99 Italian manufacturing companies, with respect to two variables: the partner variety and the innovation phase variety. Although these two variables are highly correlated (.71; p<.001), intermediate cases (i.e. companies for which the two variables are different in their value, low or high) were found among companies. As a result, four different models for open innovation were found in the practice of companies: open innovators, specialised collaborators, integrated collaborators, closed innovators. The two extreme models – open and closed - are far more diffused (in coherence with the correlation found between the two variables), so the intermediate ones need a more dedicated analysis to confirm what is emerging here, that is difficult to generalise because of the limited number of companies included in these clusters. Open and closed innovators actually emerge from this analysis as two significantly different open innovation models, especially in terms of:

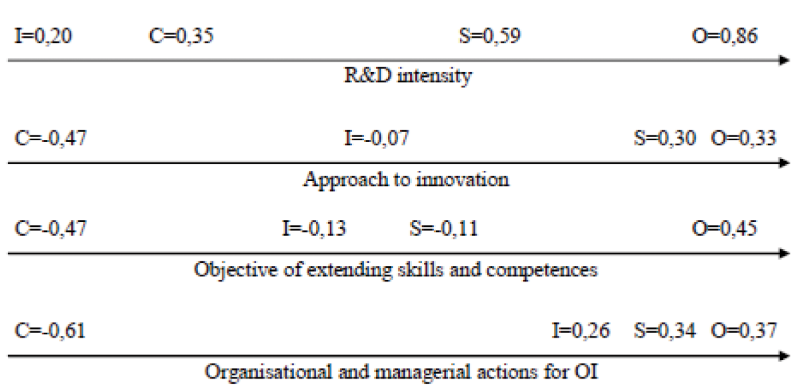
- **Approach to innovation:** open innovators are those who choose an aggressive technology and innovation strategy, in which they work to be technological leaders, to come first to the market with new products, to lead the technology evolution with superior know-how, to pursue even radical innovations. In other words, it can be argued that opening the innovation process to a wide variety of partners and all along the innovation funnel is conceived as part of an aggressive strategy;
- **R&D intensity:** open innovators invest more in R&D than closed, and this in some way confirms the difference in the strategy for the two clusters; aggressive innovators spend significantly more in R&D and, as part of their effort, they spend for opening up their innovation process. Another interesting explanation of this result refers to the need to invest in internal competences in order to be open, as the absorptive capacity of the company is critical to identify and exploit potential collaborations and exchanges with external partners. In perfect coherence with this result, there is evidence that an input-oriented organisational structure for the R&D activities, which maximises the absorptive capacity, is typical for open innovators rather than for closed;
- **Type of objectives:** in coherence with the two results above, open innovators, with respect to closed ones, mainly open their innovation process to achieve benefits related to internal competences, i.e. to deepen and integrate the knowledge base, to increase creativity and flexibility, to achieve excellence in knowledge production. On the other side, closed innovators are rather focused on reducing the costs and risks of innovation, by sharing them with external partners;

- **Organisational and managerial actions implemented to support openness:** open innovators have actually modified their organisational structure and management techniques, by introducing roles, routines and tools especially dedicated to the design, development and implementation of collaborations with external partners;
- **The results concerning the two "intermediate" models of open innovation,** i.e. specialised collaborators and integrated collaborators, are less robust, because of the limited number of companies found for these two clusters. As a consequence, only some tentative interpretation of the achieved results can be put forward. However, it is relevant to reflect on such results since they can represent the starting point for a future research aimed at verifying whether these two open innovation models can actually represent a valid alternative to open and closed models and, if so, in which specific context conditions. By making a synthesis of all results achieved for specialised and integrated collaborators, it seems that there is a sort of "continuum" in the openness of companies, in terms of the most relevant context conditions emerged in this study, as shown in Figure 3.

Figure 3 clearly shows that specialised and integrated collaborators can be really considered as "intermediate" models: the most significant variables that characterise the open and closed models, in fact, have values that are between the two extremes[4]. Integrated and specialised collaborators are thus viable options for companies that don't have a highly aggressive approach to innovation and that don't want to invest too much for opening up the innovation process. As a consequence, these companies have limited expectations in terms of benefits deriving from open innovation, but do not want to completely abandon the opportunity to access to external sources of knowledge and competencies. As an example, let's take a specialised collaborator, the electronic company cited above: the general manager wanted to spend a limited amount of resources on studying opportunities for opening the innovation process, but, at the same time, R&D managers clearly felt the need to integrate their knowledge with external contributions coming from other industries, universities, excellent research centres. As a consequence, they decided to open only the idea generation phase to a wide variety of partners.

Even in terms of performance the four models are different and a first tentative conclusion in this sense is that the degree of openness is posi-

Figure 3. Specialised and integrated collaborators between the two extreme models (C=closed innovators; I=integrated collaborators; S= specialised collaborators; O=open innovators)



tively correlated to the innovative performance: from closed innovators to open ones the level of innovative performance increases (in terms of new products and services, time to market, level of novelty, learning, costs for new products). But this does not seem to have an effect on the company's economic performance in the short term, as already discussed above. This result is in contrast with other studies (Lichtenthaler 2009), which found a positive correlation between the degree of openness (measured in terms of intensity of outbound licensing) and the economic performance (measured through ROS and ROI). In our opinion, the relation between performance and open innovation is very complex to be studied: performance is intrinsically a multi-dimensional concept (think, just to quote a well known framework, to the balanced scorecard concept), as well as the degree of openness. This can probably lead to many different measures for evaluating the relationship between openness and performance that certainly requires further in depth studies.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Our area of research is related and contributes to innovation because collaborations and networks (in particular, technological collaboration), and the proper ways to manage them, are today largely recognized as means to improve, or at least to support, firms' innovation capabilities (Chesbrough, 2006). The study is conducted by means of a survey involving 99 Italian companies operating in manufacturing industries. Different models for open innovation are found in the practice of companies: open innovators, specialized collaborators, integrated collaborators, closed innovators. The two extreme models – open and closed - are far more diffused and actually emerge as two significantly different open innovation models, in terms of approach to innovation, R&D intensity, type of objectives, organizational and managerial actions implemented to support openness. The two intermediate models - specialized and the integrated collaborators - although in need of further empirical investigation, provide evidence in support of Chesbrough's (2003b) theoretical proposition that businesses may be located along an Open Innovation Continuum.

In conclusion, the chapter introduces a new perspective that integrates both the number/typology of partners and the number/typology of phases, in order to understand if such perspective can confirm the existence of different open innovation models. Moreover, it provides useful managerial implications because it suggests that OI is not an "on/off" choice, but it can be interpreted and adopted with different degrees (Chesbrough, 2003b), consistently with the company's specific context. Thus, intermediate open innovation models (i.e. integrated and specialized collaborators) are viable options for companies that do not have a highly aggressive approach to innovation and that do not want to invest too much for opening up the innovation process. As a consequence, these companies have limited expectations in terms of benefits deriving form open innovation, but do not want to completely abandon the opportunity to access to external sources of knowledge and competencies. We suggest that intermediate models for opening the innovation process can be a first relevant topic for future research; performance of open innovation can be a second one and, finally, a third one may concern the study of open innovation models in a dynamic perspective, i.e. analyzing the path followed by companies to open their innovation process (Chiaroni et al., 2009). Adopting a dynamic perspective, the different models found in this study may be interpreted as different steps in a long term path towards open innovation: starting from a closed innovation process, companies may gradually open to a very limited set of well known partners (suppliers and customers) in a integrated collaboration model, or may decide to open only a single phase of the innovation process to a wide variety of partners with a specialized collaborator approach. Some of the cases studied in a previous work (Lazzarotti & Manzini, 2009) seem to confirm this hypothesis. A future research with longitudinal case studies can probably improve the understanding of this dynamic path to open innovation.

Finally, it should be noted that the number of respondents is still very limited. Moreover, it is studied only the relationship between some firm-specific factors and the degree of openness (defined specifically in terms of partner variety and phase variety): a wider investigation is recommendable to include more contextual factors, i.e. external/environmental ones, or more variables that can help to define the openness degree.

REFERENCES

Barbaranelli, C. (2006). *Analisi dei Dati: Tecniche Multivariate per la Ricerca Sociale*. LED Edizioni.

Brockhoff, K., & Pearson, A. (1992). Technical and marketing aggressiveness and the effectiveness of research and development. *IEEE Transactions on Engineering Management*, *39*, 318–324. doi:10.1109/17.165413

Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings and future directions. *Academy of Management Review*, *20*(2), 343–378.

Calantone, R. J., Cavusgil, S. T., & Zhao, Y. (2002). Learning orientation, firm innovation capability and firm performance. *Industrial Marketing Management*, *31*, 515–524. doi:10.1016/S0019-8501(01)00203-6

Calantone, R. J., & Stanko, M. A. (2007). Drivers of outsourced innovation: An exploratory study. *Journal of Product Innovation Management*, *24*, 230–241. doi:10.1111/j.1540-5885.2007.00247.x

Cheng, C. J., & Shiu, E. C. C. (2008). Re-innovation: The construct, measurement and validation. *Technovation*, *28*, 658–666. doi:10.1016/j.technovation.2007.08.002

Chesbrough, H. W. (2003a). *Open innovation: The new imperative for creating and profiting from technology*. Boston, MA: Harvard Business School Press.

Chesbrough, H. W. (2003b). The era of open innovation. *MIT Sloan Management Review, SPRING*, *2003*, 35–41.

Chesbrough, H. W. (2006). *Open business models: How to thrive in the new innovation landscape*. Boston, MA: Harvard Business School Press.

Chesbrough, H. W., & Crowther, A. K. (2006). Beyond high tech: Early adopters of open innovation in other industries. *R & D Management*, *36*, 229–236. doi:10.1111/j.1467-9310.2006.00428.x

Chesbrough, H. W., & Teece, D. J. (1996). When is virtual virtuous? Organizing for innovation. *Harvard Business Review*, *74*(1), 65–71.

Chiang, Y. H., & Hung, K. P. (2010). Exploring open search strategies and perceived innovation performance from the perspective of inter-organizational knowledge flows. *R&D Management*. *Special Issue Open Innovation*, *40*(3), 292–299.

Chiaroni, D., Chiesa, V., & Frattini, F. (2009). The open innovation journey: How firms dynamically implement the emerging innovation management paradigm. *Technovation*, *29*(9), 588–595.

Chiesa, V. (2001). *R&D strategy and organization, managing technical change in dynamic contexts*. London, UK: Imperial College Press.

Chiesa, V., & Manzini, R. (1998). Organising for technological collaborations: A managerial perspective. *R & D Management*, *28*(3), 199–212. doi:10.1111/1467-9310.00096

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*, 128–152. doi:10.2307/2393553

Dahlander, L., & Gann, D. (2007). *How open is innovation?* Paper presented at DRUID Summer Conference. Appropriability, Proximity, Routines and Innovation, Copenhagen, CBS, Denmark, June 18 – 20.

De Backer, L.-B. K. V., & Martinez, C. (2008). *Open innovation in a global perspective: What do existing data tell us?* (OECD Science, Technology and Industry Working Papers). OECD publishing. doi:10.1787/230073468188.

Enkel, E., Gassmann, O., & Chesbrough, H. (2009). Open R&D and open innovation: Exploring the phenomenon. *R & D Management*, *39*(4), 311–316. doi:10.1111/j.1467-9310.2009.00570.x

Enkel, E., Gassmann, O. & Chesbrough, H. (2010). The future of open innovation. *R&D Management*, special issue open innovation, *40*(3), 213-221.

European Commission. (2005). *The new SME definition. User Guide and Model Declaration*. Bruxelles, Belgium: Enterprise and Industry Publications.

Forza, C. (2002). Survey research in operations management: A process-based perspective. *International Journal of Operations & Production Management*, *22*(2), 152–194. doi:10.1108/01443570210414310

Gassmann, O. (2006). Opening up the innovation process: Towards an agenda. *R&D Management*. *Special Issue Open Innovation*, *36*, 223–228.

Gassmann, O., & Enkel, E. (2004). *Towards a theory of open innovation: Three core process archetypes*. Paper presented at R&D Management Conference, Lisbon, Portugal.

Global Business Summit. (2010). *Crescita e sfide globali*, Il Sole 24 ore. Harvard Business Review, Milano, May, 17-18.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2006). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice-Hall.

Huang, Y., Chung, H., & Lin, C. (2009). R&D sourcing strategies: Determinants and consequences. *Technovation*, *29*, 155–169. doi:10.1016/j.technovation.2008.08.002

Huston, L., & Sakkab, N. (2006). Connect and develop: Inside Procter and Gamble's new model for innovation. *Harvard Business Review*, *85*, 58–66.

Keupp, M. M., & Gassmann, O. (2009). Determinants and Archetype users of open innovation. *R & D Management*, *39*(4), 331–341. doi:10.1111/j.1467-9310.2009.00563.x

Kolk, A., & Püümann, K. (2008). Co-development of open innovation strategy and dynamic capabilities as a source of corporate growth. *Journal of Economic Literatur*em, (n.d), 73–83.

Laursen, K., & Salter, A. (2006). Open for innovation: The role of openness in explaining innovation performance among U. K. manufacturing firms. *Strategic Management Journal*, *27*(2), 131–150. doi:10.1002/smj.507

Lazzarotti, V., & Manzini, R. (2009). Different modes of open innovation: A theoretical framework and an empirical study. *Journal of Innovation Management*, *13*(4), 1–22.

Lichtenthaler, U. (2008). Open innovation in practice: An analysis of strategic approaches to technology transactions. *IEEE Transactions on Engineering Management*, *55*(1), 148–157. doi:10.1109/TEM.2007.912932

Lichtenthaler, U. (2009). Outbound open innovation and its effect on firm performance: Examining environmental influences. *R & D Management*, *39*(4), 317–330. doi:10.1111/j.1467-9310.2009.00561.x

Lichtenthaler, U., & Ernst, H. (2009). Opening up the innovation process: The role of technology aggressiveness. *R & D Management*, *39*(1), 317–329. doi:10.1111/j.1467-9310.2009.00561.x

Linder, J. C. (2004). Transformational outsourcing. *MIT Sloan Management Review*, *45*(2), 52–58.

Lundvall, B. A. (1992). *National systems of innovation: Towards a theory of innovation and interactive learning*. London, UK: Pinter.

Lynch, C. F. (2004). Why outsource? *Supply Chain Management Review*, *8*(7), 44–51.

MacCormack, A., & Iansiti, M. (2009). Intellectual property, architecture, and the management of technological transition: Evidence from Microsoft Corporation. *Journal of Product Innovation Management*, *26*, 248–263. doi:10.1111/j.1540-5885.2009.00656.x

MINE SURVEY TOOLS 2.1 (n. d.). *École Politechnique Montreal* (Survey), Social Sciences and Humanities Research Council of Canada, Industrial Research Institute. Available on www.minesurvey.polymtl.ca

Osservatorio Unioncamere Brevetti e Marchi Report. (2008). *Osservatorio Unioncamere Brevetti e Marchi Report*. ISBN: 886077067X ISBN-13: 9788860770677.

Perrons, R. K., Richards, M. G., & Platts, K. (2005). What the hare can teach the tortoise about make-buy strategies for radical innovations. *Management Decision*, *43*, 670–690. doi:10.1108/00251740510597707

Piachaud, B. (2005). Outsourcing technology. *Research Technology Management*, *48*(3), 40–46.

Pisano, G. P., & Verganti, R. (2008). Which kind of collaboration is right for you? *Harvard Business Review*, (December): 1–9.

Porter, M. E. (1980). *Competitive strategy: Techniques for analyzing industries and competitors*. New York, NY: Free Press.

Sakkab, N. (2002). Connect & develop complements research & develop at P&G. *Research-Technology Management*, *45*(2), 38–45.

Slowinski, G., & Zerby, K. W. (2008). Protecting IP in collaborative research. *Research and Technology Management, November-December*, 58-65.

Sofka, W., & Grimpe, C. (2008). *Managing search strategies for open innovation – The role of environmental munificence as well as internal and external R&D*. Discussion Paper No. 08-075, Centre for European economic research.

Sofka, W., & Grimpe, C. (2010). Specialized search and innovation performance – Evidence across Europe. *R&D Management, Special Issue Open Innovation, 40*(3).

Szulanski, G. (1996). Exploiting internal stickiness: Impediments to the transfer of best practice. *Strategic Management Journal, Winter Special Issue, 17*, 27–43.

Trott, P., & Hartmann, D. (2009). Why "open innovation" is old wine in new bottles. *International Journal of Innovation Management*, *13*(3), 715–736. doi:10.1142/S1363919609002509

van de Vrande, V., deJong, J., Vanhaverbekec, W., & deRochemont, M. (2008). Open innovation in SMEs: Trends, motives and management challenges. *Technovation*, *29*, 423–437. doi:10.1016/j.technovation.2008.10.001

van de Vrande, V., Lemmens, C., & Vanhaverbeke, W. (2006). Choosing governance modes for external technology sourcing. *R & D Management*, *36*(3), 347–363. doi:10.1111/j.1467-9310.2006.00434.x

vandeMeer, H. (2007). Open innovation — The Dutch treat: Challenges in thinking in business models. *Creativity and Innovation Management, 6*(2), 192–202.

Vanhaverbeke, W. (2006). The interorganisational context of open innovation. In H. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), *Open Innovation: Researching a new paradigm* (pp. 205–219). Oxford, UK: Oxford University Press.

von Hippel, E. (1988). *The sources of innovation*. New York, NY: Oxford University Press.

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, re-conceptualization and extension. *Academy of Management Review*, *27*, 185–203.

KEY TERMS AND DEFINITIONS

Absorptive Capacity: The ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends (Cohen & Levinthal, 1990).

Closed Innovation Model: In closed innovation model a company generates, develops and commercializes its own ideas. This philosophy of self-reliance dominated the R&D operations of many leading industrial corporations for most of the 20th century (Chesbrough, 2003b).

Inbound Open Innovation (*Outside-In Process*): Enriching the company's own knowledge base through the integration of suppliers, customers and external knowledge sourcing (Gassmann & Enkel, 2004).

Open Innovation: Firms rely increasingly on external sources of innovation, which means that ideas, resources and individuals flow in and out of organizations (Chesbrough, 2003a).

Open Innovation Continuum: Innovation models which the firms follow are not exclusively open or closed, but rather show varying degrees of openness. Between the two pure models - open or closed - there are many shades of grey and the dichotomy between open vs. closed is artificial (Chesbrough, 2003b; Dahlaner & Gann, 2007).

Outbound Open Innovation (*Inside-Out Process*): Earning profits by bringing ideas to market, selling IP and multiplying technology by transferring ideas to the outside environment (Gassmann & Enkel, 2004).

Technology Aggressiveness: Firms with aggressive technology strategies constantly developing new technologies that are superior to the technologies of their competitors. R&D activities of firms with aggressive technology strategies are usually highly specialized. Based on this specialization, the firms tend to focus on radical rather than incremental innovations (Brockhoff, & Pearson, 1992; Lichtenthaler & Ernst, 2009).

ENDNOTES

1. The other two items are: the specialization of the company's R&D activities and the high importance of development activities (the D part of R&D) relative to the firm's overall R&D activities (Lichtenthaler and Ernst, 2009).
2. We applied the criteria suggested by the EU (European Commission, 2005) for classifying firms on the basis of their size. Specifically, an autonomous company is classified as: (i) small, if the number of workers is < 50 and turnover is <= 10 million € or the annual balance sheet total is <= 10 million € (ii) medium, if the number of workers is between 50 and 250 and turnover is between 10 and 50 million € or the annual balance sheet total is between 10 and 43 million €; (iii) large, if the number of workers is > 250 and revenues > 50 million € or the annual balance sheet total is > 43 million €.
3. Data about correlations between intensity of collaboration with each typology of partner and each typology of phase are not reported in this chapter but they are available upon request.
4. Only R&D intensity is lowest in the case of integrated collaborators, but this is coherent with the fact that closed innovators need to invest a lot in R&D, since they have to develop internally all tangible and intangible resources needed for innovation.

APPENDIX

1. Partner variety: in the last five years you have collaborated with a wide variety of external actors
2. Phase variety: in the last five years you have collaborated on a wide variety of phases.
3. Intensity of collaboration with (each) partner: in the last five years you have collaborated very strongly with the following partner (University and Research centres, Technical and Scientific Service Companies, Governmental institutions, Customers, Suppliers, Competitors, Firms operating in different sectors of activity).
4. Intensity of collaboration on phase: in the last five years you have collaborated very strongly on the following phases (Idea generation, Experimentation, Engineering, Manufacturing set up, Commercialization).
5. Objectives of collaboration
    a. aims to extend skills, competences and creativity:
        i. Enlarge the company's competence base
        ii. Increase the flexibility of the internal organization
        iii. Stimulate creativity and idea generation capability
    b. aims to share risks and costs:
        i. Reduce or share the risks of innovation
        ii. Reduce or share the costs of innovation
6. Approach to innovation (technology aggressiveness)
    a. Investing for technological leadership
    b. Aggressive acquiring new business areas by means of innovation
    c. Influencing the industry structure and rules by means of products characteristics
    d. Trying to recruit the best researchers and experts available on the market
    e. Giving emphasis on radical rather than incremental innovation
7. Organizational and managerial actions for open innovation
    a. Top management is committed towards the maximization of the collaborations results
    b. Personal relationship of the R&D manager are exploited to start technological collaborations
    c. For each collaboration, there is a "champion" acting as a facilitator for the collaboration success
    d. The company formally evaluates the objectives and risks of the collaboration
    e. The company analyses and selects the potential partners with a formal and explicit process
8. Innovative performance
    a. The company's competence base was enlarged
    b. The average development costs of new products/processes was reduced
    c. The time to market of new products / processes was reduced
    d. The level of innovativeness of new products / processes was improved
    e. Sales volume and market acceptance of new products was improved

This work was previously published in Technological, Managerial and Organizational Core Competencies: Dynamic Innovation and Sustainable Development, edited by Farley Simon Nobre, David Walker and Robert J. Harris, pp. 167-190, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Section 6
Managerial Impact

This section presents contemporary coverage of the social implications of Industrial Engineering, more specifically related to the corporate and managerial utilization of information sharing technologies and applications, and how these technologies can be extrapolated to be used in Industrial Engineering. Core ideas such as service delivery, gender evaluation, public participation, and other determinants that affect the intention to adopt technological innovations in Industrial Engineering are discussed. Equally as crucial, chapters within this section discuss how leaders can utilize Industrial Engineering applications to get the best outcomes from their shareholders and their customers.

Chapter 70
Offshoring Process:
A Comparative Investigation of Danish and Japanese Manufacturing Companies

Dmitrij Slepniov
Aalborg University, Denmark

Brian Vejrum Wæhrens
Aalborg University, Denmark

Hiroshi Katayama
Waseda University, Japan

ABSTRACT

The purpose of this chapter is to contribute to the knowledge on how production offshoring and international operations management vary across cultural contexts. The chapter attempts to shed light on how companies approach the process of offshoring in different cultural contexts. In order to achieve this objective, the authors employ a qualitative methodology and compare three Danish and three Japanese manufacturing companies. On the basis of this comparative investigation, the authors find that the parent companies from both contexts employ offshoring as a remedy for the challenges of globalisation. Yet there are clear differences in how offshoring is conducted in Denmark and Japan. The main differences are outlined in a framework and explained employing cultural variables. The findings lead to a number of propositions suggesting that the process of offshoring is not simply a uniform technical-rational calculation of the most efficient organisation of activities across national borders, but it is rather specific to the parent companies' national contexts.

INTRODUCTION

The IT-revolution, market liberalisation, market integration, innovative production technologies and numerous other factors related to globalisation have enabled many tasks to be performed at a distance. In order to utilise the best locations for discrete value-added activities, companies configure these activities on a global scale. As a result, growing numbers of production companies resemble dynamic, complex globally dispersed and interconnected webs of internal and external

DOI: 10.4018/978-1-4666-1945-6.ch070

relationships, rather than static and traditional co-located production systems.

In this chapter, the global dispersion of value-added activities is referred to as offshoring. This process arises through substituting overseas activities for domestic activities; as a rule, it also involves a re-configuration of the existing global operations set-up (Farrell, 2004; Pyndt & Pedersen, 2006; Lewin & Couto, 2007). Offshoring is closely interrelated with the management of global supply chain relationships. The nature of offshoring suggests that it inevitably leads to more complex and varying relationships amongst entities in globally stretched configurations of operations.

The topic of offshoring has received extensive attention in the literature. Since the 1960s, the international business (IB) literature has dealt with various aspects of why multinational corporations (MNCs) exist and how they organize the relationship between headquarters and international subsidiaries (e.g. Dunning, 1977; Bartlett & Ghoshal, 1986). The strategy literature has applied the lenses of transaction cost economics (TCE) (Williamson, 1975), resource-based view (RBV) (Barney, 1991) and core competencies (Hamel & Prahalad, 1990) to the analysis. The operations management (OM) literature has discussed the changing roles of manufacturing (e.g., Voss, 2005; Riis et al., 2007) and newly emerging manufacturing systems organised as global manufacturing networks (Ferdows, 1997; Shi & Gregory, 1998).

In spite of their variety, the theoretical approaches to the topic of offshoring have been predominantly rooted in the inherent economic logic. The factors and variables of offshoring have been treated as uniform and universal for a very broad spectrum of leading firms from the industrial triad of North America, Western Europe and Japan. This perspective is based on a basic assumption that, in order to remain competitive, these lead firms have to move parts of their operations activities to the low-cost regions of non-Japan Asia, Latin America and Eastern Europe (e.g., Quinn, 1999; Mol et al., 2005; McIvor, 2005; Kakabadse & Kakabadse, 2003). As vital and valid that this assumption might be, the issue of how the process of offshoring varies across the cultural context of these developed countries remains largely unexplored.

The research objective of this chapter is, therefore, to add to the existing literature by exploring whether and how offshoring process patterns depend on the cultural context of a dispatching organisation. We refer to the dispatching organisation as the parent company, located in the domestic base from which an offshoring initiative originates. The receiving organisation, on the other hand, is defined as the wholly owned subsidiary or third party facility in a foreign country to which operations activities are being relocated.

We recognize the importance of the receiving context and its fit with the dispatching context. However, this chapter primarily focuses on the dispatching organisation and the role of its cultural context. The underlying premise is that the initial offshoring decision, as well as the process that follows it, is influenced more by the dispatching organisation than by its differentiated partners and affiliates on the receiving side. The support for this premise can also be found in the literature. For example, Buckley (2009) argues that spatial distribution strategies make the role of parent organisations more important in the new global set-up than in conventional firms.

The objectives of the study are achieved by conducting a comparative investigation of Danish and Japanese companies. We identify a number of differences in the companies' offshoring journeys and discuss the role of the parent companies' cultural context which, among other factors, has influenced a particular approach to offshoring.

We pursue these arguments in the rest of the chapter. The following section introduces the concepts and theories employed in the investigation. Next, we introduce recent studies that outline the patterns of offshoring in Danish and Japanese companies. This is followed by six case studies from these countries. Finally, we present the analysis and discussion of offshoring processes in companies from the Danish and Japanese cultural contexts.

THEORETICAL BACKGROUND

Offshoring: Modes, Drivers and Impediments

Over the past decades, the world has become flat, as 'flat as the screen on which a business leader can host a meeting of his whole supply chain' (Friedman, 2005:7). In these 'flat' conditions operations are becoming increasingly dispersed geographically; growing numbers of companies of various sizes are turning to offshoring, in an attempt to increase their competitiveness on a global basis (Farrell, 2004; Mol et al., 2005). However, offshoring has been one of the most hotly debated research topics in recent years; there is no consensus in the literature about what is even meant by the term 'offshoring'. One reason for this is the complexity and multi-dimensionality of this phenomenon. It deals with international competitiveness and growing interdependence of firms from developed and developing countries; a large part of the offshoring debate touches upon the global labour market and its effects on jobs, wages and skills in various parts of the world; there is also a focus on how economics activities are organised across firm and country boundaries.

This chapter is concerned with the latter aspects of offshoring. As has been mentioned above, we define offshoring as the global dispersion of value-added activities. But even this definition of offshoring is not straightforward, and its meaning has changed over time. In its early years, offshoring as the global spread of production was examined mostly from the 'in-house' and foreign direct investment (FDI) perspectives. The first studies that contributed to understanding operations-related offshoring issues originating in the international business (IB) literature stream (e.g., Dunning, 1977; Bartlett & Ghoshal, 1986). Four different types of FDI were identified: market-seeking, efficiency-seeking, resource-seeking and strategic asset-seeking. These are related to location advantages of the national subsidiaries of the multinational enterprise (MNE), which was one of the primary research units in the IB literature. Back then, offshoring was viewed as a mean for international expansion through FDI and entering new markets by establishing new national subsidiaries.

Although the above perspective is conceptually clear and might have been adequate in the 1960-1980s, it hardly captures the complexity of how operations are likely to be organised today. In this new organisation, national subsidiary companies have been replaced by a series of discrete and dispersed value-added activities (Birkinshaw, 2003). Parent companies configure these activities on a global scale. From this perspective, offshoring can be defined as the relocation of an operations activity or an entire manufacturing facility to a foreign country (Cavusgil et al., 2008). We adopt this view on offshoring in the remainder of the chapter.

The relocation of activities can take place in two ways: (1) captive and (2) offshore outsourcing. The captive way of offshoring refers to the process of relocating company's activities overseas without giving up their ownership and direct control. In other words, captive offshoring occurs on an in-house or 'intrafirm' basis. Offshore outsourcing, on the other hand, can be viewed as a complete or partial discontinuation of in-house domestic or in-house international activities and, thus, refers to externally supplied or 'outsourced' activities.

Obviously, captive offshoring and offshore outsourcing do not represent the full spectrum of offshoring possibilities. They rather show two extreme scenarios, which can be further differentiated, in terms of both their contractual and location implications. Kotabe & Murray (2004) argue that in terms of contractual arrangements offshore outsourcing can further be split into arm's length relationships and strategic partnerships. From a location point of view, companies can relocate activities either to geographically distant countries or 'nearshore'. The Economist (2005) defines nearshoring as the business of moving activities 'to countries that are quite cheap and very close rather than very cheap and far away'.

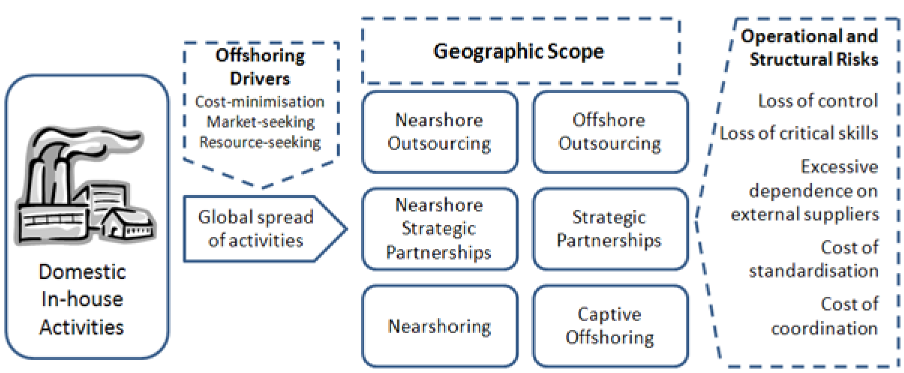
By configuring value-adding activities across the world, companies reap various benefits of offshoring. These benefits include getting access to cheaper factors of production, being close to important markets and getting access to knowledge and other strategic resources (Ferdows, 1997; Doh, 2005). If offshoring takes place in the offshore outsourcing way, the parent company may also benefit from the economies of specialisation and focus on the core competencies (Hamel & Prahalad, 1990). However, on the other hand, a number of factors also impede the trend towards offshoring. First of all, the cost of coordinating globally dispersed value-adding activities may be high (Kotabe & Murray, 2004). Next, offshoring of some activities, proprietary activities in particular, may lead to operational risks (e.g., quality-related) and structural risks (e.g., decreasing brand value) (Aron & Singh, 2005). In the case of offshore outsourcing, both the operational and structural risks are particularly severe. They may potentially lead to the loss of control over activities, loss of critical skills, declining innovation capability of the parent company, and excessive dependence on external suppliers (e.g., Kotabe, 1992; McIvor, 2005).

The choice of the offshoring mode also depends on which value-added activities are to be transferred overseas. While highly proprietary activities (e.g., R&D, design, branding) have the overall tendency to be concentrated at home, standardised activities (e.g., production) are more conductive to being geographically dispersed using the offshore outsourcing mode (Cavusgil et al., 2008).

Figure 1 builds on the reviewed literature and outlines the major offshoring modes, drivers and impediments briefly discussed in this section. The factors and variables of offshoring articulated here will be employed in the following parts of the chapter.

The literature also identifies a number of theoretical competing explanations for why and how companies offshore. Institutional theories (e.g., Greenwood & Hinings, 1996) emphasize congruence with shifting industry norms and shared logics. The transaction cost economics (TCE) approach (e.g., Williamson, 1975) focuses on intra-firm vs. inter-firm transactions costs and achieving the optimum level of internalization of activities. The resource-based view (RBV) (e.g., Barney, 1991) stresses the importance of firm-core-related resources and capabilities in sustaining a firm's competitive advantage. These explanations focus on the rational search for an optimal

Figure 1.



fit between performance objectives and operations resources. They also typically portray companies from developed countries as a homogeneous group 'renting' the cheaper capacity from overseas partners. However, we should not rush into making fast conclusions about uniformity and universality of the location and make-or-buy decisions of companies from the traditional industrial centres of North America, Western Europe and Japan.

Relating Offshoring Processes to Cultural Context

Westeney (2003), discussing how Japanese companies differ from US and European companies, outlines three sets of factors determining potential differences: 1) system-level factors related to history, dominant management ideologies, and investment regimes; 2) home country effects related to generic patterns of strategy institutionalised in the domestic business system; 3) firm-level factors related to a company's international experience.

The firm level factors are dynamic and tend to converge across companies from different countries over time. However, the inherent economic logic that any firm follows does not account for all aspects of its behaviour, and differing patterns still occur. Some of them may well be attributed to system-level factors, such as government policies. They may play an important role in shaping firms' offshoring patterns. However, the full assessment of the role of such factors can only be conducted working from a base in political economy and is beyond the scope of this chapter. Instead, we choose to focus on company's country of origin impacts in terms of national culture. Our reasoning for this choice is based on the argument that introducing the cultural dimension of offshoring will allow for probing non-rational and county-specific aspects of offshoring patterns, as they develop through a complex and evolutionary interplay of historical events and human agents.

The notion that the country of origin matters in various aspects of global business is generally accepted in the IB literature (e.g. Kogut & Singh, 1988). However, in the operations and supply chain management literature, the impact of national culture on offshoring has been somewhat peripheral to the discussion. A study by Voss and Blackmon (1998) represents one of the rare exceptions. It addresses cultural differences as a factor contributing to different operations strategy orientations between Japan and the West. Several other studies (e.g., Pagell et al., 2005; Ang & Inkepen, 2008) also call for employing cultural variables in dealing with various aspects of international operations management.

However, what is culture, and how can the issue of the possible cultural relativity of offshoring practices be approached? In the second half of the 20th century, many authors attempted to define culture (e.g. Kroeber & Parsons, 1958; Hofstede, 1983; Schein, 2004). In this chapter, we adopt Hofstede's (1983) cross-disciplinary definition of culture as the collective mental programming of people within a national culture, with consequences for beliefs, attitudes, and skills. Hofstede's (1983, 1984, 2001) dimensional approach to cross-cultural comparison includes five dimensions of national cultural variation: Collectivism vs. Individualism, Large vs. Small Power Distance, Strong vs. Weak Uncertainty Avoidance, Masculinity vs. Femininity, Long vs. Short-Term Orientation.

The approach for arriving at these dimensions, as well as the way of measuring them, is not immune to criticisms. McSweeney (2002), one of Hofstede's harshest critics, questions several aspects of Hofstede's approach. First, concerns are raised about the representativeness of the sample used in Hofstede's study, which was conducted over thirty years ago. The data were collected using a survey of professionals only from one location in each country and from employees of only one company (IBM). Next, according to McSweeney (2002), because the results of the survey show considerable variation within samples in individual countries, we can only resort to using average tendencies when comparing nations. Finally, exposing the limitations of the functionalist approach, the critics cast doubts about whether the five dimensions capture all the complexities of a national culture which is not always territorially bound. Hofstede (2001) counters many of these criticisms, but also admits that his 'world-scale dimensions cannot do justice to the profound meanings of local practices in the countries' (Hofstede, 2001:73).

Nevertheless, since the 1980s, when four out of five dimensions were introduced, the ideas of Hofstede have become an integral part of any major study dealing with culture. The number of societies covered in the Hofstede's research – more than 50 – includes similar as well as different ones. This number is large enough to describe the research as geocentric, rather than just ethnocentric or polycentric. Compared to other multidimensional classifications of culture we considered (e.g. Inkeles & Levinson, 1997), Hofstede's research also stands out as grasping the notion of culture in a manageable and suitable manner for international business and operations management. These factors determined the use of Hofstede's indices in our analysis. In the discussion part of the chapter, we link each of the five cultural dimensions to patterns of offshoring observed in the literature and practice in the contexts of two developed countries: Denmark and Japan.

Before we proceed, it is important to note that, despite our focus on national culture, we acknowledge that in a business context, culture can develop at different levels – national, organisational, professional. Our deliberate focus on national culture is based on the argument articulated in numerous studies on cross-cultural management that it is the national culture in which an organisation is embedded that has been found to be the main influence (Browaeys & Price, 2008).

In the following section, we compare offshoring trajectories in companies from the two countries. The question is: what link, if any, can be established between the offshoring strategies of companies and their cultural context?

COMPARISON OF OFFSHORING IN DENMARK AND JAPAN

Patterns Highlighted in the Literature

The contexts of Denmark and Japan offer a good opportunity for a comparative analysis. Apart from differences in size, both economies have some important similarities. Both countries enjoy the same high level of industrialisation and economic development, both have high labour costs, and the manufacturing sectors in both countries have traditionally been *first forte* of the their economies. So what is already known about offshoring patterns in Denmark and Japan?

First of all, neither of the two countries represents an exception to the growth in scale and scope of the offshoring phenomenon. Kimura and Ando (2005) suggest that Japanese firms have been increasingly capitalizing on the opportunities arising from offshoring of production activities. The geographic targets of Japanese offshoring initiatives as well as the logic behind these initiatives, has shifted over time. In 1980, North America was the primary destination for Japanese FDI. Access to market, tariff avoidance and yen-dollar exchange rates were among the major factors in the location of 'transplants' of Japanese companies to the U.S. in the period of 1970-1990 (Schroeder et al., 1992).

However, over time, the market entry motivation behind offshoring has been decreasing due to lower trade barriers. Katayama et al. (1999) point out that, since the 1990s, growing numbers of world-class Japanese industries have been increasingly shifting their operations to low-cost Asian countries, in order to deal with the challenges of maintaining competitiveness through getting access to cheaper labour and resources. During this period, Asia became one of the prime locations for the inflow of manufacturing investments by Japanese companies (Westney, 2003).

Hijzen et al. (2006) find that, despite widespread concerns that offshoring leads to a hollowing-out effect and job losses, the Japanese firms engaged in offshoring actually tend to strengthen their economic activities in Japan both in terms of output and employment. Westney (2003) finds that in many Japanese companies which were actively extending their production networks abroad in the 1990s, the export ratio fell very little or in some cases even rose considerably.

There is evidence to suggest that, in shifting their activities overseas, Japanese firms tend to chose captive offshoring as opposed to offshore outsourcing. Tomiura (2005) reports that only a small fraction, less than 3%, of Japanese manufacturing firms, outsource their production from independent suppliers (*dukuritsu kaisha*). According to The Economist (2007), relative lack of enthusiasm for engaging in the offshore outsourcing mode is partially due to the fact that in Japan, outsourcing decisions are seen as uncomfortable and undermining corporate pride.

Benito et al. (2002), studying internationalisation patterns of Scandinavian companies, Danish among them, find that companies from this region demonstrate a very high propensity to internationalise their operations. Statistics Denmark (2008) reported that 19% of all Danish companies (with 50 or more employees) sourced internationally. The most frequent destinations for offshoring in Danish companies were the new EU member states, as well as low-cost countries in South East Asia (Statistics Denmark, 2008). The same study found that captive offshoring and offshore outsourcing were equally important types of offshoring by Danish companies. Some research points out that Danish firms tend to reduce their domestic operations especially with regard to labour-intensive jobs. For example, according to Pyndt and Pedersen (2006), 80% of all textile-related jobs in Denmark were lost as a consequence of offshoring of labour-intensive activities to low-cost regions. In general, in the latest wave of offshoring, efficiency seeking is seen by Danish companies as the main driver for shifting activities abroad (Statistics Denmark, 2008)

Jenster et al.'s (2006) study analyzed offshoring trends in Europe. It found that, in terms of jobs loses to offshoring, Denmark is two to three times more severely affected than the European Union average. However, it is important to note that synergetic and dynamic effects of offshoring also result in an increase in knowledge-intensive job creation in Denmark (Jenster et al., 2006; Statistics Denmark, 2008). Although these findings demonstrate ambiguity about the impact of offshoring on the Danish labour market, one thing seems certain - Danish companies take a rather aggressive approach to offshoring and downsize their domestic base.

Table 1 summarizes the major features of offshoring patterns, distilled from the reviewed studies focusing on offshoring by Danish and Japanese companies.

To enrich the empirical base for our investigation, in the following we present cases of three Japanese and three Danish companies. The case descriptions together with the material discussed above, form the basis for the analysis which follows the case studies section.

Case Examples from Denmark and Japan

Methodology

The case study, one of several strategies for doing qualitative research, has been chosen for this investigation for several reasons. First, case studies can describe, enlighten and explain real-life phenomena that are too complex for strategies requiring tightly structured designs or pre-specified data sets. On the other hand, case studies do not have to be ethnographic or participatory action research based. They are 'defined by interest in individual cases, not by methods of enquiry'

Table 1. Differing patterns of offshoring in Danish and Japanese companies

	Offshoring Drivers	*Offshoring Mode*	*Offshoring Effects*
Japan	Market-seeking Cost-minimisation Resource-seeking Currency exchange rate fluctuations	Mixed, but predominantly captive and strategic partnerships	Domestic core intact or enhanced
Denmark	Cost-minimisation Market-seeking Resource-seeking	Mixed, but with high degree of offshore outsourcing	Domestic core downsized (especially in labour-intensive activities)

(Stake, 2003:134). Second, instrumentally, the case study strategy is well equipped for furthering understanding of particular contemporary issues or concepts which have not been deeply investigated so far (Eisenhardt, 1989; Voss et al. 2002; Yin, 2009). Third, according to Yin (2009), case studies are generally preferred for answering 'how' and 'why' questions about a contemporary phenomenon, over events in which the investigator has little or no control.

Our research objective satisfies all these conditions (i.e., the type of questions, phenomenon under investigation and controllability) and thus is well suited for the case study strategy. The study is concerned with how the process of offshoring varies across cultural contexts. As we established, studies focusing on cultural variables in the context of offshoring process are scarce and the cultural relativity of offshoring is still at the understanding and discovery stage. As far as the controllability issue is concerned, case studies again emerge as the preferred method for this study. On the one hand, we had sufficient access to the actual phenomenon through company visits and interviews with managers involved in offshoring initiatives. On the other hand, the control we had over the events in the cases was not sufficient for applying other methods, such as a participatory action research.

Six exploratory case studies (three representing the Danish context and three – the Japanese context) were selected randomly on the basis of possessing two key attributes: 1) a long-standing involvement in offshoring initiatives; 2) a strong need for closeness to markets. The latter criterion was introduced in order to control for the market-proximity factor on the firm level. Considerations about the access to potential data, including commitment of interviewees, and availability of documents, were also taken into account.

The offshoring process in all six companies started prior to our involvement in the cases. Therefore, many events relevant to the study had to be captured in retrospect. The data collection was conducted following a formal research protocol. However, the nature of case study research is rather iterative and unfolding (Strauss, 1987; Pettigrew, 1990). It was almost impossible to routinize the data collection because the specific information was often not readily available. In these instances, alternative leads had to be pursued.

The data collected for the case study was organised in a database. Table 2 outlines two main elements of the database indicating which data were collected and how.

Documents and archival records, as the sources of evidence, formed a significant amount of the case studies database. They had several strengths: could be reviewed repeatedly, provided broad coverage of time and events, and contained exact spelling of names and references. All the collected documents and records were extensively studied prior and after the site visits and interviews.

Interviews constituted another source of evidence. The interviews were conducted as in-depth

Table 2. Value of the five dimensions for Denmark and Japan

Country	Individualism/ Collectivism		Power Distance		Uncertainty Avoidance		Masculinity/ Femininity		Long/Short Term Or.	
	Index	Rank	Index	Rank	Index	Rank	Index	Rank	Index	Rank
Denmark	74	9	18	51	23	51	16	50	46	10
Japan	46	22-23	54	33	92	7	95	1	80	4

Source: Hofstede (2001)

guided conversations. Respondents were asked to give detailed accounts of their companies' offshoring initiatives. The focus was put on chronology of events, major drivers, location decisions, decision making authority, and areas of organisation affected by offshoring. Moreover, the respondents were asked about how the initiative had affected their job and its task environment. The interview data had its challenges and weaknesses. This source of evidence is not immune to biases, poor recollection of events and inaccuracies. To minimize the impact of these weaknesses, the interview data were triangulated with the documents and records we collected.

All these data combined were used to build the case narratives presented below. Within-case and more general cross-case analyses enabled us to identify patterns of offshoring, which we compare with those identified in the literature and later discussed in relation to the pre-selected cultural dimensions.

Case A: Cautious Offshoring Expansion with Retaining Manufacturing Competencies

Company A is a Japanese company, which develops and produces bearings as well as other friction control products, for a wide variety of industries. Currently it has approximately 1500 employees worldwide. 1100 of them are based in Japan, where the company's headquarters and its five manufacturing sites are located. The company also operates four wholly owned manufacturing sites overseas in the USA, China, Thailand and Czech Republic. It also has a wide network of differentiated affiliates, mainly involved in sales and distribution.

The company's production offshoring journey started in 1986 with the establishment of manufacturing facilities in the USA. In 1988, this was followed by establishing a subsidiary in Germany, where production of some components for the automotive industry started in 1991. These investments were made in order to be close to key customers and to reduce the negative impact of the strong yen on the company's export prices.

The offshoring initiative got a big boost during the last decade. However, the motivation for offshoring has changed. The recent offshoring moves were largely cost-minimisation driven. In this period, several manufacturing sites were opened in the low-cost locations. In 1998, the company started its manufacturing activities in Shanghai, China. In 2002, a sales office and manufacturing facilities were established in Thailand. In 2005, the company's European manufacturing base was transfered from Germany to the Czech Republic. The same year, activities in China were expanded and a new site in Suzhou was established.

Implications of the latest offshoring initiatives for the domestic base were very limited. To some extent, this can be explained by a low level of connectivity between plants. In such fragmented system, the domestic base remains shielded from the changes happening in the company's global operations set-up. However, it is also part of the company's strategy to preserve a strong domestic base, which is perceived as paramount to maintaining a long-term competitive advantage. In 2002-

2008 at all its Japanese facilities, the company implemented a new production system, based on the Toyota Production System (TPS) principles. This was done to ensure that manufacturing in Japan remained competitive and was retained as one of the company's core competencies. Thus, the company approaches its overseas facilities as a complement, rather than a substitute, to its domestic production. It remains persistently committed to retaining its manufacturing competences in-house in its domestic base. The same approach is taken towards the development function. Although some development activities take place in the company's facilities in the USA, they are tightly controlled from Japan. There are no plans to replicate or intensify technology development in other parts of the world.

Case B: Emphasising Ownership and Consolidating Mission Critical Activities at Home

Case B is a Japanese company specialising in measurement, control and automation technologies and equipment. The company's main production facilities are located in Japan, the US and China, with workforce of approximately 5500 employees. Its overseas affiliates and subsidiaries, focusing mainly on sales and customer support, are located in 11 other countries. The company prioritises full ownership of its subsidiaries, at home as well as overseas, and sees this as a central means for reinforcing its business sustainability.

Traditionally, the company's international links were limited to collaboration with an American partner, which had an equity stake in the case company for several decades. However, in 1990, this equity stake was considerably reduced. In the 1990s, the company started a gradual expansion of its presence in Asia. In spite of unfavourable economic conditions in the region at the time, the company opened sales and customer support offices in Thailand, Indonesia, Philippines, Singapore, Korea, Malaysia, and Taiwan. In 2001, the company also established a manufacturing base in Europe as a result of a 12-year journey, which started in 1989 with the opening of a sales office in Belgium.

The facilities in Europe were established primarily with the objective to improve access to the market there. However, in addition to market-seeking, in the 1990s, the company also started looking for low-cost manufacturing locations. In 1992, the company established its first joint-venture in China, where a new manufacturing plant was opened in 1995. Over a number of years, these facilities, which currently are 100% owned by the company, played a number of roles. First, the Chinese site was used as a server for the south-east Asian market, demanding less sophisticated lower-cost products. Second, by supplying precision die-cast components to the Japanese plants, the facilities in China played a crucial role in the company's steps towards strengthening the cost-competitiveness of its production at the domestic base in Japan. In 2008, similar roles were assigned to newly opened wholly owned facilities in Vietnam. As was the case with other overseas facilities, the operations in Vietnam were a result of a carefully planned 3 year initiative. It started with opening a representative office in the country and later was followed by a gradually increasing commitment to this location.

In the end of the 1990s, a new International Business Division was set up in the company's headquarters in Tokyo. The division was established in order to centralise all the affiliates' activities and provide a global overview of overseas operations. The role of the division increased considerably in 2002, when the company terminated its long-term equity alliance with the American partner after gradually reducing the partner's stake in the business for a number of years. Since then, the Tokyo headquarters has become responsible for the coordination and development of sales and operations worldwide. Previously, this activity was largely the prerogative of the American partner.

The company sees itself as a technology-oriented manufacturer. In 2007, its R&D function, previously scattered across its operating base, was consolidated at the company's new technology centre near Tokyo. Concentrating research, product development and engineering at the domestic base was a strategic step, providing 'one-stop' solutions in a close proximity to the main Japanese manufacturing sites.

Case C: Preserving Strategic Competences in the Domestic Base

Case C is a Tokyo-based manufacturer of industrial instruments and control systems. It currently employs approximately 5000 people. Close to 50% of the company's sales are generated outside its home market, in 32 countries from all over the world. This has been achieved through consistent long-term efforts in global business expansion.

As early as the 1970s, the first production facilities outside Japan were established in Singapore. In 1986, the company started operations in China. In the following years, the number of new overseas production plants gradually increased. In 2007, the company was operating 14 plants in 9 different countries including high-cost countries, such as the US, the Netherlands, Singapore and Germany, as well as low-cost countries, such as China, Korea, Brazil, Indonesia, and India. Its presence in the high-cost countries was crucial for having better visibility and access to strategic markets, whereas, the plants in low-cost countries represented the company's attempt to enhance production capabilities in low-cost destinations and ensure a low-cost supply to the Japanese facilities.

The strategy of the company was not only to keep as much production in Japan as possible, but also to preserve its core development capabilities in the domestic base. The first R&D centres were opened outside Japan, in Singapore and India. These new centres, in total employing 85 people, played a more peripheral support role to the main R&D function, with 1000 engineers located in Japan. The centre in Singapore was supposed to develop better links with customers, particularly outside Japan. The idea of the new centre was also to practice a style of communication and thought-process that engineers based in Japan could gradually learn from. The mission of the team of engineers in India was somewhat similar to that of the Singapore office, i.e. to improve the link with overseas customers. The engineers in India were trained to speak and even to think in Japanese, so they could communicate effectively with colleagues in Japan. The company started training its Indian team 10 years before the office was open, so they could contribute more to product development projects led by engineers in Japan.

Case D: Aggressive Offshoring Cycle - from Production to Competence Centre Overseas

Company D is a Danish company, which develops and produces acoustics and mechanics components for the telecommunication industry. It is known for its high quality standards and is on the forefront of innovation in the industry. The company has approximately 4000 employees worldwide, of which 75 are based at its headquarters in Denmark. It operates R&D and sales facilities in the USA and the Netherlands, as well as five wholly owned production sites in Poland, China, and Vietnam.

In the past 5-7 years, the company experienced a period of enormous intensity of offshoring. In 2001-2002, company's production in the Netherlands was relocated to Poland. In 2005, the bulk of production was moved from Poland to Vietnam. The Polish site was transformed into pre-production and R&D centre.

In the period of 1999-2006, the company transitioned from lead production in Denmark to lead production in China, through a number of intermediate steps. In 1999, a customer service centre was established in China to serve the rising Asian market. In 2000, full scale production was transferred from Denmark to China. This move was mainly based on cost considerations. The market was in a slump, and the company believed that the transfer would lead to an increased focus on cost. A secondary driver, but increasingly important, was emergence of the local market in China. At that time, all product and process development, along with some ramp-up activities, were retained in Denmark. However, by 2006, after realizing that having two parallel sites was counterproductive, all production activities were relocated from Denmark to China. The production set-up in China was very different from the Danish one. In particular, production processes in China were more manual work based, while only quality sensitive processes were automated.

Moving production to China affected all other functions of the company. In 2005, a small R&D department was established in Shanghai, to support ramp-up activities and indigenous development activities in China. In 2006, it was decided to give the Chinese production site the role of competence centre for production, thus recognising that production competence was no longer primarily based in Denmark. The proximity to suppliers of key components played an important role in this decision, and a new purchasing department was established in China. It took over all operational and tactical tasks from the Danish parent company, which preserved only the authority over the strategic purchasing decisions.

In the end of 2006, a new production site in Vietnam was opened. The main motivation for establishing the new facilities was related to lower labour costs. Production at the new site primarily complements production at the Chinese facilities, where lead production continues to take place.

Case E: Outsourcing Initiative with an Ambitious Scale and Timeline

Company E is a Danish producer of products characterised by high seasonality and exposure to fashion trend fluctuations. The company employs approximately 4000 employees worldwide. It had traditionally been committed to its operations in Denmark. For a number of years, it had also operated its own production sites in Switzerland and the United States. However, in 2004, its perspective on global operations changed considerably. Its emphasis switched from in-house production to outsourcing in low-cost countries: Mexico, Hungary and the Czech Republic.

The transition was in line with the company's financial direction towards reduction of fixed assets and achieving an 'asset-light' state. The initiative was supposed to take up to three years, in which 80% of operations capacity had to be transferred to the company's external partners. The ambitious scale and the timeline of the initiative posed a number of challenges for the company. For example, high volatility in demand, due to seasonal fluctuations and changing trends, prevented the company from building sufficient buffer stock for the transition period. It posed a serious challenge of maintaining on-going production while transferring production to external partner sites and coordinating their activities. Next, difficulties were encountered with constantly transferring know-how to external partners. Approximately 60-70% of the company's annual turnover was generated through the continuous launch of new products. This required very close collaboration between product development and the supply chain, with an uninterrupted flow of information and know-how between the two. In the new outsourcing reality, this flow had been substantially hampered because the customer of the development process had changed from internal to geographically and cognitively distant external production units. However, these challenges did not stop the analytical group, involving a number of middle-level managers in charge of the initiative, from engaging the company in the quickly escalating offshoring initiative.

The focus on managing the process was increased though the introduction of a dedicated monthly process responsible for monitoring and coordinating the production network based on the capacities, roles and responsibilities assigned to various facilities. The company had been heavily relying on this process in coordinating its partners' activities with its own lead plant activities remaining in Denmark. Although the process proved to be successful in managing the network during the transition period, in the beginning of 2008 it also revealed its limitations, when the company announced its plan to 'backsource' or bring back in-house one plant run by its partners.

Case F: Shift in Paradigm - from Manufacturing to System Integration

Company F is an SME offering high quality woven textiles to industrial customers. Originally, it mastered a broad variety of production processes in-house. In the late 1990-s, the accelerating velocity of competition and transformations in the external environment of the company revealed the need for changes in its strategic activities. During this period, the company started outsourcing some of its activities and later also began to buy finished fabrics from external suppliers.

Outsourcing was a major strategic decision affecting the entire organisation and was vividly compared with an octopus reaching into every nook and cranny of the business. Having been a manufacturing company in the past, today it primarily manages a network of suppliers and serves as a systems integrator with close contact to customers. This involves not only managing daily operations and logistics, but also monitoring and negotiating with all actors in the value chain. For example, the logistics manager maintains personal contact with the vendors of wool in New Zealand. The company buys the wool and calls on full scale spinning and weaving factories abroad to process the wool to specifications.

At factories in one of the Baltic countries, weaving takes place on machinery owned by the company. The machines were previously located in Denmark and are currently leased by the foreign supplier. The demand for high quality has led the Danish company to maintain and further develop its production engineering competencies when the spinning and weaving operations were outsourced. The partner's plant outside Denmark was essentially an offshore full-scale operation focused on cost and efficiency, while product and process development, laboratory and prototype production resided in the Danish headquarters. After the offshoring journey of the company started in 2000, some quality sensitive operations, such as dying, softening and washing, were carried out by the company's own main factory in Denmark. However, in 2006, the decision was made to transfer them abroad, as well. The radically shifted the role of the Danish facilities from manufacturing to innovation, systems integration and supply chain management. Supplier relationship management became one of the focus areas of the company. Because the strategic partners of the company are not contractually bound, the relational approach plays the key role in how the company manages its operations network.

The subsequent downstream processes, such as sewing and upholstery, had traditionally been carried out by customers and were of their concern. But now, in line with developing the new role as a systems integrator, the company also offers the service of finding suppliers for these processes. The company is not likely to undertake production itself, but is more likely to further build up sourcing and negotiation skills, e.g. to offer prototyping and ramp-up at suppliers' sites.

DISCUSSION

Patterns Observed in the Cases

The main aim of this discussion is to generate a framework, leading to a number of propositions relating offshoring to dimensions of national culture. However, first we summarise the patterns of offshoring observed in the six case studies. The cases presented above can be analysed in relation to the offshoring process variables identified in the theory section of this chapter. Table 3 summarises the characteristics of the cases in terms of their production offshoring mode, drivers, impediments, implications, pace and geographic spread. The strategic trajectories found in the six cases correspond to the general trends identified in the literature.

The data point to a number of similarities and differences between the Japanese and Danish companies. Significant similarities can be found in: 1) their motivation for initiating offshoring projects and 2) the geographic spread of offshoring. Differences, on the other hand, are visible across all other dimensions of offshoring, including mode, pace and factors hampering further expansion of offshoring in these companies. The impacts of offshoring on domestic bases also differ in the Japanese and Danish cases. In the Danish cases, the process of transferring production activities overseas led to the downsizing or complete closure of the domestic production facilities. In contrast, the domestic plants in the Japanese cases remained almost intact and even enhanced their strategic role, with respect to foreign plants.

Many factors may explain these offshoring trajectories and singling out one factor may be difficult. Therefore, we apply the unexplained variance approach (Poole et al., 2000). We ask ourselves, what aspects of the offshoring processes in the cases cannot be explained by the system-level and firm-level factors? In these instances, culture may become an important explanatory factor.

Both at the system and firm levels, the case companies have similar goals for engaging in the offshoring of activities. At a macro level, there is a country size differential, but in terms of policies and the economic development level both countries are similar and represent high-wage and knowledge-intensive economies. At the firm level, all six companies represent industries with a relatively high labour intensity and a propensity towards offshoring. Although we cannot fully isolate the cultural factor, these similarities at the macro and firm levels justify the exploration of the role of national culture as a potentially important variable affecting their offshoring developments.

Table 3. The offshoring process characteristics in the cases

	Japan			Denmark		
	Case A	**Case B**	**Case C**	**Case D**	**Case E**	**Case F**
Offshoring impacts	Domestic production and R&D bases intact	Domestic R&D enhanced; Production base intact	Domestic production and R&D bases intact	Domestic production phased out; R&D downsized	Domestic production downsized	Domestic production phased out
Mode	Captive	Captive	Captive	Captive	Offshore Outsourcing	Offshore Outsourcing
Geographic spread	Mix of nearshoring and distant locations	Mix of nearshoring and distant locations	Mix of nearshoring and distant locations	Mix of nearshoring and distant locations	Mix of neashoring and distant locations	Neashoring
Drivers	Cost Market Currency exchange rates	Cost Market Currency exchange rates	Cost Market Currency exchange rates	Cost Market	Cost Market	Cost Market
Impediments	Loss of control; loss of core business skills	Loss of control; loss of core business skills	Loss of control; loss of core business skills	Loss of core business skills; loss of control	Loss of core business skills; coordination cost	Loss of control; coordination cost
Pace	Slow and incremental	Slow and incremental	Slow and incremental	Fast and radical	Fast and radical	Fast and radical

The cultural lens may help us to reveal that the offshoring process is more than just an anonymous response to the emerging challenges and opportunities of the global market, but rather that it is strongly accorded with a unique set of cultural characteristics of the parent company.

Cultural Context Impacts on Offshoring

Hofstede's (1983, 2001) five dimensions of national culture provide a useful framework for accounting for the effects of culture in the offshoring process. As Table 4 indicates, Denmark's and Japan's national cultural contexts differ substantially across all five dimensions. Denmark is characterised as a feminine, highly individualistic society with low power distance, weak uncertainty avoidance and a short term orientation. In contrast, Japan is a more collectivist society, with high power distance, a strong tendency towards uncertainty avoidance, and traditional masculine values permeating the society having one of the highest long term orientation indices.

In the following discussion, these features are related to offshoring patterns identified in the two contexts. The first dimension of national culture is concerned with how one individual is related to other individuals in a society. As can be seen from Tables 4 and 5, Japan is much more collectivistic than Denmark. According to Hofstede (1983), in a collectivistic society, the degree of interdependence between organisations and their members is much higher than in an individualist society. Individuals in a collectivistic society can expect other individuals and organisations to look after them, in exchange for loyalty and belongingness, while in an individualistic society, the relationship between employees and employers is, to a large extent, perceived as a business transaction in a labour market. We can relate this dimension to the implications of offshoring for domestic bases in the Danish and Japanese cases. In Denmark, offshoring led to the closure or downsizing of domestic facilities, while, in Japan, it had a very limited impact on the domestic base in the companies. We can assume that the management offshoring decisions in the Japanese cases were made taking into account possible negative implications of offshoring for employees of the domestic plants. In the Japanese companies, an aggressive approach, dictated by market logic only, would imminently clash with the society's collectivistic orientation and harmony seeking. Such approach might have also been interpreted as an unacceptable loss of face, but it was justified in the Danish cases.

The second dimension of national culture is power distance. We can again point to a difference between the two countries. Japan has a higher power distance than Denmark. This dimension can be seen as complementary to the first dimension, with a negative correlation between the two, i.e. a high individualism index is related a low power distance in the society and vice-versa. This explains why hierarchical order and hierarchical organisational structures are more common in Japan than in Denmark. Relating the power distance

Table 4. Interpretation of the values of the five dimensions of culture for Denmark and Japan (based on Hofstede (2001))

Country	Individualism/ Collectivism	Power Distance	Uncertainty Avoidance	Masculinity/ Femininity	Long/Short Term Or.
Denmark	Individualistic Society	Short Power Distance	Weak Uncertainty Avoidance	Feminine Society	Short-Term Orientation
Japan	Collectivistic Society	Large Power Distance	Strong Uncertainty Avoidance	Masculine Society	Long-Term Orientation

dimension of national culture to the offshoring trajectories in the cases, we can build on Mulder's (1977) hypothesis that powerful individuals strive to maintain power distance and the existing power structure. The management of the Japanese cases strived for preserving a high power distance and existing hierarchical structure. In contrast, in the Danish cases, this factor did not play a significant role in preventing the closure and downsizing of domestic facilities. Furthermore, in the Danish cases, middle management enjoyed extended autonomy. For example, in Case F, a number of direction-determining in-process decisions were made at the middle management level. On the other hand, the Japanese cases were much more top management driven.

The third national cultural dimension is uncertainty avoidance, which addresses how a society deals with uncertainty and ambiguity. Here, again, Denmark and Japan score very differently. Denmark has weak uncertainty avoidance, while Japan has strong uncertainty avoidance. Although uncertainty avoidance should not be confused with risk avoidance, a low rating in uncertainty avoidance is an indication that a nation is generally more willing to take risks and tolerate new ideas that deviate from well established norms. The uncertainty avoidance dimension is especially important in explaining the differences in approaches to offshoring in the context of the two countries. The case companies' decisions were clearly affected by the level of tolerable uncertainty. For the Danish case companies, offshoring all or larger part of their domestic manufacturing (and in Case D, spreading the R&D function globally) to a cognitively and geographically distant low-cost countries, took a much higher risk than the Japanese companies. The latter exercised a much more cautious approach to offshoring, by keeping the production and R&D capacity and competencies domestically in-house or offshoring only to cognitively and geographically close low-cost countries of South-East Asia. Certain Japanese companies (Case A and C) could not avoid opening R&D centres overseas, in order to remain responsive to foreign markets they were operating in. Nevertheless, they approached this task very differently from the Danish companies. The Japanese cases took years to establish even peripheral support centres, while in the Danish cases, the process was much more compressed in time and the scope of activities was more far reaching.

The fourth dimension of national culture is related to how masculine or feminine a society is. In masculine societies, values related to aggressiveness and striving for success are common for the whole society or, as Hofstede (1983) puts it, the values include the importance of 'big is beautiful', while a more feminine society adheres to the 'small is beautiful' concept. Referring to Tables 4 and 5, we see that Japan and Denmark represent extremes on the scale, Denmark being a feminine society, while the masculine values prevail in Japan. These cultural norms also affected the management actions underlying the trajectory of the offshoring process in the case companies. Factors such as corporate pride played a particularly important role in guiding the offshoring decisions in the Japanese cases. On the other hand, in the Danish cases the particular mental framing of manufacturing activities played a different role and, to a large extent, served as a driver of offshoring.

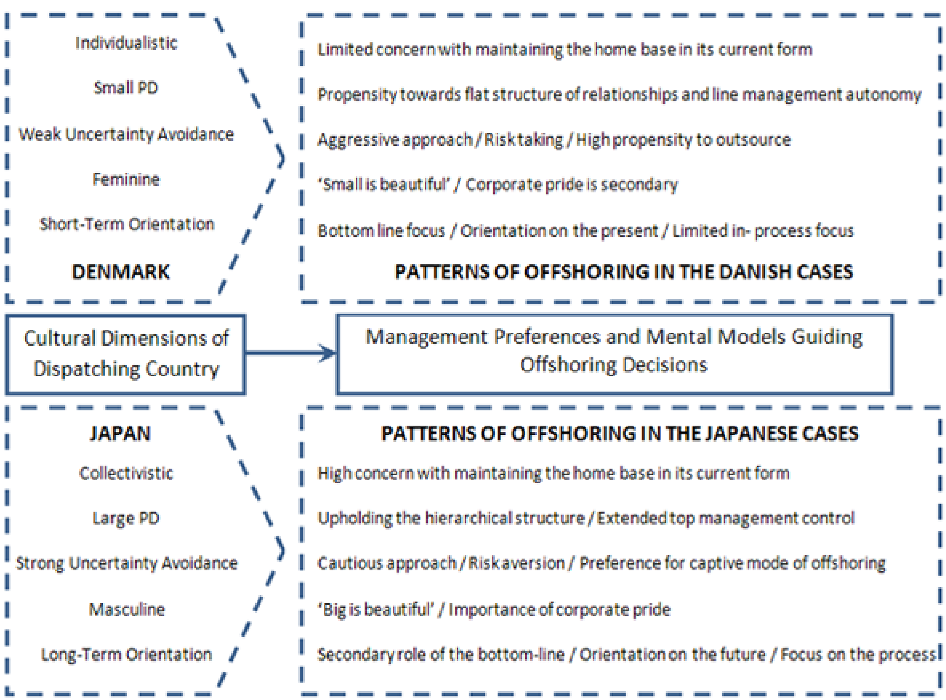
The fifth dimension of culture is long- versus short-term orientation. As its name suggests, this dimension is related to the choice of focus for people's efforts – the future or the present. As discussed by Hofstede (2001), the scores on this dimension index show that Asian countries score the highest, and Western countries are on the low-medium end. Denmark and Japan are not exceptions to this. Denmark represents the cluster with a short-term orientation, while Japan belongs in the cluster with a long-term orientation. Businesses in a long-term oriented culture are expected not to have a focus on immediate results, while in a short-term oriented culture the focus on the bottom line is a major concern of management (Hofstede, 2001; Mamman & Saffu, 1998). However, this view of time is not fully conclusive. For example, the findings of Voss and Blackmon (1998) suggest that the Western view of time is primarily linear and short-term, while the Japanese perspective is polychronic, i.e. focusing on both the long-term and short-term simultaneously. The role of this dimension is also evident in the cases used in our study. In the Danish cases, representing short-term orientation society, the offshoring initiatives are compressed in time, and management decisions seem to follow the logic of seeking quick economic results with minimum possible costs. The decision making process in the Danish cases is based on more technical rational analysis deriving from the most immediate economic reality. This is illustrated by the swift and frequent change of the strategic roles of plants in all the three Danish cases. In the Japanese cases, on the contrary, the process is spread out in time; it links short-term activities and long-term goals. It does not disregard, but rather tolerates, the extra resources required for preserving manufacturing competences in the domestic base.

The main points of the discussion presented above are summarised in Figure 2. The framework links the five cultural dimensions with the preferences and mental 'programs' of the management teams in the case companies. These preferences guide offshoring decisions and, ultimately, the distinguishing trajectories of offshoring.

This framework represents an attempt to take a more in-depth look into the process of offshoring and how this process may depend on the dispatching company's cultural context. It can be distilled into five propositions:

Proposition 1: *Companies from a cultural context dominated by collectivistic values are likely to adopt a more cautious and community responsive approach to offshoring than companies from a cultural context dominated by individualistic values,*

Figure 2.



Proposition 2*: In companies from cultural contexts dominated by low power distance the middle managers play a more prominent role in offshoring initiatives than in companies from cultural contexts dominated by a high power distance,*

Proposition 3*: In companies from cultural contexts dominated by weak uncertainty avoidance, the propensity to adopt an aggressive approach to offshorin,g involving risk and uncertainty, is higher than in companies from cultural contexts dominated by strong uncertainty avoidance,*

Proposition 4*: In companies from cultural contexts dominated by masculine values, the importance of corporate pride is more prominent than in companies from cultural contexts dominated by feminine values,*

Proposition 5*: In companies from cultural contexts dominated by a short-term orientation, the focus on short-term financial results drives offshoring processes to a larger extent than in companies from cultural contexts dominated by a long-term orientation.*

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Our main findings show how certain developments in offshoring processes and international operations management practices can be explained by differences in the dispatching companies' cultural contexts. We, by no means, claim that, based on these results, we can predict an accurate scenario for the offshoring process for every country. However, we believe that the findings and the ensuing propositions may be useful in informing our expectations about the propensity of companies from a particular context towards a particular offshoring process trajectory.

Although there is no 'one size fits all' approach to offshoring, it is clearly culturally dependent. The chapter demonstrates that, despite the offshoring phenomenon often being discussed in technical rational terms, it is not a universal process; it is, rather, subject to strong cultural influence. On the basis of the literature review and investigation of a multiple-case studies strategy, we find that, on the one hand, there are many similarities between Danish and Japanese companies. These similarities are especially visible in the realm of offshoring motivational factors (cost-minimisation and market-seeking). On the other hand, the findings of this chapter indicate that there are substantial differences between Danish and Japanese companies in: 1) how they approach offshoring in terms of mode and pace; 2) how they perceive strategic risks resulting from offshoring; and 3) how the process influences the domestic bases of the companies. We find that cultural variables affect the management viewpoints which guide their offshoring decisions. These findings support the findings of Pagell et al. (2005) about the validity of culture as an explanatory variable in operations management research. Furthermore, they also contribute to understanding of how the dimensions of culture affect specific underlying mechanisms of the offshoring process.

However, this study has a number of limitations. Rather than providing definititive answers, it develops a number of propositions and raises the need for further research on cultural variables and their influence on the offshoring process. Although the cultural approach may have sufficient explanatory power for an analytical comparison between Japanese and Danish companies, the cultural variables used in the study may not have a substantial explanatory power for intra-regional comparisons. Another limitation is that, although national culture is a valid approach to the offshoring phenomenon, there are a number of other influencing factors. Future research is needed to formulate specific propositions relating other factors to the offshoring process and cultural variables. These other factors may include the location of plants (where a firm offshores to matters to how it does it), resource-based advantages, the capital structure and size of firms.

Another limitation of the study is related to the case approach employed in the investigation. Potential biases of this strategy may occur due to selective memory of the respondents. In this study, this was offset by triangulating the interview data with related corporate documentation of the process. Next, while ensuring contextual rootedness, the case methodology offers limited generalisability (Eisenhardt, 1989). However, as examples from other studies suggest, the findings are nonetheless indicative of the respective Japanese and Danish approaches to offshoring. Moreover, the limited generalisability of the study does not undermine its findings. They still remain important for a number of reasons.

The aim of this investigation was to add to the knowledge about the offshoring phenomenon, i.e. what Woodside and Wilson (2003) call the principal objective of case study research. The findings of the chapter foster a better understanding of the process of offshoring and cultural variables as a factor significantly affecting it. The findings also can be used by managers who deal with global work transfers and challenges associated with these transfers.

REFERENCES

Ang, S., & Inkpen, A. C. (2008). Cultural intelligence and offshore outsourcing success: A framework of firm-level intercultural capability. *Decision Sciences*, *39*(3), 337–358. doi:10.1111/j.1540-5915.2008.00195.x

Aron, R., & Singh, J. V. (2005). Getting offshoring right. *Harvard Business Review*, *83*(12), 135–143.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120. doi:10.1177/014920639101700108

Bartlett, C. A., & Ghoshal, S. (1986). Tap your subsidiaries for global reach. *Harvard Business Review*, *64*(6), 87–94.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *Management Information Systems Quarterly*, *11*(3), 369–386. doi:10.2307/248684

Benito, G., Larimo, J., Narula, R., & Pedersen, T. (2002). Multinational enterprises from small economies: The internationalization patterns of large companies from Denmark, Finland and Norway. *International Studies of Management and Organization*, *32*(1), 57–78.

Bettis, R. A., Bradley, S. P., & Hamel, G. (1992). Outsourcing and industrial decline. *The Academy of Management Executive*, *6*(1), 7–22.

Birkinshaw, J. (2003). Strategy management in MNE subsidiaries. In Rugman, A., & Brewer, T. L. (Eds.), *The Oxford handbook of international business* (pp. 380–402). Oxford, UK: Oxford University Press.

Browaeys, M.-J., & Roger, P. (2008). *Understanding Cross-Cultural management*. Upper Saddle River, NJ: Prentice Hall.

Buckley, P. J. (2009). Internalisation thinking: From the multinational enterprise to the global factory. *International Business Review*, *18*, 224–235. doi:10.1016/j.ibusrev.2009.01.006

Cavusgil, S. T., Knight, G., & Riesenberg, J. R. (2008). *International Business: Strategy, Management, and the New Realities*. Upper Saddle River, NJ: Prentice Hall.

Doh, J. P. (2005). Offshore outrourcing: implications for international business and strategic management theory and practice. *Journal of Management Studies*, *42*(3), 695–704. doi:10.1111/j.1467-6486.2005.00515.x

Dunning, J. H. (1977). Trade, location of economic activity and the MNE: a search for an eclectic approach. In Ohlin, B., Hesselborn, P. O., & Wijkman, P. M. (Eds.), *The international allocation of economic activity* (pp. 395–418). London: Macmillan.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550. doi:10.2307/258557

Farrell, D. (2004). Beyond offshoring: Assess your company's global potential. *Harvard Business Review*, *82*(12), 82–90.

Ferdows, K. (1997). Made in the world: The global spread of production. *Production and Operations Management*, *6*(2), 102–109. doi:10.1111/j.1937-5956.1997.tb00418.x

Friedman, T. (2005). *The World is Flat: A Brief History of the Twenty-First Century*. New York: Allen Lane.

Greenwood, R., & Hinings, C. R. (1996). Understanding radical organizational change: bridging together the old and the new institutionalism. *Academy of Management Review*, *21*(4), 1022–1054. doi:10.2307/259163

Hamel, G., & Prahalad, C. K. (1990). The core competence of the corporation. *Harvard Business Review*, *68*(3), 79–91.

Hayes, R., Pisano, G., Upton, D., & Wheelwright, S. (2005). *Operations, Strategy and Technology: Pursuing the Competitive Edge. Danvers*. John Wiley & Sons.

Hijzen, A., Inui, T., & Todo, Y. (2006). The effects of multinational production on domestic performance: Evidence from Japanese firms. *RIETI Discussion Papers*. Series 07-E-006. Retrieved August 1st, 2008, from http://www.rieti.go.jp/jp/publications/dp/07e006.pdf.

Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of International Business Studies*, *14*(2), 75–89. doi:10.1057/palgrave.jibs.8490867

Hofstede, G. (1984). Cultural dimensions in management and planning. *Asia Pacific Journal of Management*, *1*(2), 81–99. doi:10.1007/BF01733682

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nnations*. Thousand Oaks, CA: Sage Publications.

Inkeles, A., & Levinson, D. J. (1997). National character: The study of modal personality and socio cultural systems. In Inkeles, A. (Ed.), *National Character: A Socio-Social Perspective*. New Brunswick: Transaction.

Jensen, P. D., Kirkegaard, J. F., & Laugesen, N. S. (2006). Offshoring in Europe – Evidence of a two-way street from Denmark. *Institute for International Economics Working Papers*, WP 06-3. Retrieved February 29, 2008, from http://www.petersoninstitute.org/publications/ wp/wp06-3.pdf.

Kakabadse, A., & Kakabadse, N. (2003). Outsourcing best practice: transformational and transactional considerations. *Knowledge and Process Management*, *10*(1), 60–71. doi:10.1002/kpm.161

Katayama, H., Sirichan, K., & Hiraki, S. (1999). Japanese offshore manufacturing in Thailand: Its difficulties and future prospects. *Integrated Manufacturing Systems*, *10*(4), 210–220. doi:10.1108/09576069910280459

Kimura, F., & Ando, M. (2005). Two-dimensional fragmentation in East Asia: conceptual framework and empirics. *International Review of Economics & Finance*, *1*(3), 317–348. doi:10.1016/j.iref.2004.12.005

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, *19*(3), 411–433. doi:10.1057/palgrave.jibs.8490394

Kotabe, M., & Murray, J. Y. (2004). Global sourcing strategy and sustainable competitive advantage. *Industrial Marketing Management*, *33*, 7–14. doi:10.1016/j.indmarman.2003.08.004

Kroeber, A. L., & Parsons, T. (1958). The concepts of culture and of social system. *American Sociological Review*, *23*, 582–583.

Lewin, A. Y., & Couto, V. (2007). *Next Generation Offshoring: The Globalization of Innovation*. Chicago: Booz Allen Hamilton.

Mamman, A., & Saffu, K. (1998). Short-terminism, control, quick-fix and bottom line. *Journal of Managerial Psychology*, *13*, 291–308. doi:10.1108/02683949810219864

McIvor, R. (2005). *The Outsourcing Process: Strategies for Evaluation and Management*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511543425

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith – a failure of analysis. *Human Relations*, *55*(1), 89–118.

Mol, M. J., Rob, J. M., van Tulder, R., & Beije, P. R. (2005). Antecedents and performance consequences of international outsourcing. *International Business Review*, *14*, 599–617. doi:10.1016/j.ibusrev.2005.05.004

Mulder, M. (1977). *The Daily Power Game*. Leiden: Martinus Nijhoff.

Pagell, M., Katz, J. P., & Sheu, C. (2005). The importance of national culture in operations management research. *International Journal of Operations & Production Management*, *25*(4), 371–394. doi:10.1108/01443570510585552

Pettigrew, A. M. (1997). What is a processual analysis? *Strategic Management Journal*, *13*(4), 5–16.

Poole, M., Van de Ven, A. H., Dooley, K., & Holmes, M. E. (2000). *Organizational Change and Innovation Processes*. Oxford: Oxford University Press.

Pyndt, J., & Pedersen, T. (2006). *Managing Global Offshoring Strategies*. Copenhagen: CBS Press.

Quinn, J. B. (1999). Strategic outsourcing: leveraging knowledge capabilities. *Sloan Management Review*, *40*(4), 9–21.

Riis, J. O., Johansen, J., Waehrens, B. V., & Englyst, L. (2007). Strategic roles of manufacturing. *Journal of Manufacturing Technology Management*, *18*(8), 933–948. doi:10.1108/17410380710828262

Schein, E. H. (2004). *Organizational Culture and Leadership*. San Francisco: Jossey-Bass.

Schroeder, R. G., Sakakibara, S., Flynn, E. J., & Flynn, B. B. (1992). Japanese plants in the U.S.: How good are they? *Business Horizons*, *35*(4), 66–72.

Shi, Y., & Gregory, M. (1998). International manufacturing networks – to develop global competitive capabilities. *Journal of Operations Management*, *16*(2-3), 195–214. doi:10.1016/S0272-6963(97)00038-7

Stake, R. E. (2003). Case Studies. In Denzin, N. K., & Lincoln, Y. S. (Eds.), *Strategies of Qualitative Inquiry* (pp. 134–165). Thousand Oaks, CA: Sage Publications.

Statistics Denmark. (2008). *International Sourcing: Moving Business Functions Abroad*. Copenhagen: Statistics Denmark.

The Economist. (2005). *The rise of nearshoring*. December 1st, 61-63.

The Economist. (2007). *No country is an island*. A special report on business in Japan. December 1st, 15-16.

Tomiura, E. (2005). Foreign outsourcing and firm-level characteristics: Evidence from Japanese manufacturers. *Journal of the Japanese and International Economies*, *19*, 255–271. doi:10.1016/j.jjie.2005.03.007

Voss, C. (2005). Alternative paradigms for manufacturing strategy. *International Journal of Operations & Production Management*, *25*(12), 1211–1222. doi:10.1108/01443570510633611

Voss, C., & Blackmon, K. (1998). Differences in manufacturing strategy decisions between Japanese and Western manufacturing plants: The role of strategic time orientation. *Journal of Operations Management*, *16*(2-3), 147–158. doi:10.1016/S0272-6963(97)00034-X

Voss, C., Tsikriktis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, *22*(2), 195–219. doi:10.1108/01443570210414329

Westney, D. E. (2003). Japan. In Rugman, A., & Brewer, T. L. (Eds.), *The Oxford handbook of international business* (pp. 623–652). Oxford: Oxford University Press.

Williamson, O. E. (1975). *Markets and Hierarchies*. New York: Free Press.

Woodside, A. G., & Wilson, E. J. (2003). Case study research methods for theory building. *Journal of Business and Industrial Marketing*, *18*(6/7), 493–508. doi:10.1108/08858620310492374

Yin, R. K. (2009). *Case Study Research - Design and Methods*. Thousand Oaks: Sage Publications.

ADDITIONAL READING

Benn, I., & Pearcy, J. (2002). *Strategic Outsourcing: Exploring the Skills of Third Parties*. London: Hodder & Stoughton.

Brown, D., & Wilson, S. (2005). *The Black Book of Outsourcing: How to Manage the Changes, Challenges, and Opportunities*. New York: John Wiley & Sons.

Buckley, P. J., & Casson, M. C. (1976). *The Future of the Multinational Enterprise*. London: Macmillan.

Coe, N. M., Dicken, P., & Hess, M. (2008). Introduction: Global production networks - debates and challenges. *Journal of Economic Geography*, *8*(3), 267–269. doi:10.1093/jeg/lbn006

Farrell, D. (2006). *Offshoring: Understanding the Emerging Global Labor Market*. Boston: Harvard Business School Press.

Ferdows, K. (1997). Made in the world: The global spread of production. *Production and Operations Management*, *6*(2), 102–109. doi:10.1111/j.1937-5956.1997.tb00418.x

Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of International Political Economy*, *12*(1), 78–104. doi:10.1080/09692290500049805

Grant, E. B., & Gregory, M. J. (1997). Adapting manufacturing processes for international transfer. *International Journal of Operations & Production Management*, *17*(10), 994–1005. doi:10.1108/01443579710176997

Hamilton, G. G., & Biggart, N. W. (1988). Market, culture, and authority: A comparative analysis of management and organization in the Far East. *American Journal of Sociology*, *94*, S52–S94. doi:10.1086/228942

Heywood, B. J. (2001). *The Outsourcing Dilemma: the Search for Competitiveness*. London: Prentice Hall.

Hira, R., & Hira, A. (2005). *Outsourcing America: What's Behind our National Crisis and How We Can Reclaim American Jobs*. New York: Amacom.

Hofstede, G. (2002). Dimensions do not exist: A reply to Brendam McSweeney. *Human Relations*, *55*(11), 1355–1361.

Jahns, C., Hartmann, E., & Bals, L. (2006). Offshoring: Dimensions and diffusion of a new business concept. *Journal of Purchasing and Supply Management*, *12*(4), 218–231. doi:10.1016/j.pursup.2006.10.001

Jenster, P., Pedersen, H. S., Plackett, P., & Hussey, D. (2005). *Outsourcing - Insourcing: Can Vendors Make Money from the New Relationship Opportunities?* Chichester: John Wiley & Sons.

Johnson, M. (1997). Outsourcing. In *Brief*. Oxford, UK: Butterworth-Heinemann.

Kakabadze, A., & Kakabadze, N. (2003). Outsourcing best practice: Transformational and transactional considerations. *Knowledge and Process Management*, *10*(1), 60–71. doi:10.1002/kpm.161

Kshetri, N. (2007). Institutional factors affecting offshore business process and information technology outsourcing. *Journal of International Management*, *13*(1), 38–56. doi:10.1016/j.intman.2006.05.003

Lonsdale, C., & Cox, A. (2000). The historical development of outsourcing: The latest fad? *Industrial Management & Data Systems*, *100*(9), 444–450. doi:10.1108/02635570010358384

Lyles, A., & Salk, J. E. (2007). Knowledge acquisition from foreign parents in international joint ventures: An empirical examination in the Hungarian context. *Journal of International Business Studies*, *38*, 3–18. doi:10.1057/palgrave.jibs.8400243

Mayer, K. J., & Argyres, N. S. (2004). Learning to contract: Evidence from the personal computer industry. *Organization Science*, *15*(4), 394–410. doi:10.1287/orsc.1040.0074

McIvor, R. (2009). How the transaction cost and resource-based theories of the firm inform outsourcing evaluation. *Journal of Operations Management*, *2*(1), 45–63. doi:10.1016/j.jom.2008.03.004

Momme, J. (2002). Framework for outsourcing manufacturing: Strategic and operational implications. *Computers in Industry*, *49*(1), 59–75. doi:10.1016/S0166-3615(02)00059-3

Porter, M. E. (1986). Changing patterns of international competition. *California Management Review*, *28*(2), 9–40.

Preston, S. (2004). Lost in migration: Offshore need not mean outsourced. *Strategy and Leadership*, *32*(6), 32–36. doi:10.1108/10878570410568893

Quinn, J. B. (1999). Strategic Outsourcing: Leveraging knowledge capabilities. *Sloan Management Review*, *40*(4), 9–21.

Schniederjans, M. J., Schniederjans, A. M., & Schniederjans, D. G. (2005). *Outsourcing and Insourcing in an International Context*. London: M.E.Sharpe.

Smircich, L., & Stubbart, C. (1985). Strategic management in an enacted world. *Academy of Management Review*, *10*(4), 724–736. doi:10.2307/258041

Williamson, D. (2002). Forward from a critique of Hofstede's model of national culture. *Human Relations*, *55*(11), 1373–1395.

KEY TERMS AND DEFINITIONS

Captive Offshoring: The type of offshoring which occurs on an 'intrafirm' (in-house) basis.

Culture: The collective mental programming with consequences for beliefs, attitudes, and skills.

Dispatching Organisation: A parent company located in the domestic base where an offshoring initiative originates from.

Offshore Outsourcing: The type of offshoring which occurs on an 'inter-firm' basis.

Offshoring Process Trajectory: A set of variables, such as location and organisational boundary decisions, which are assumed by an offshoring initiative over time under the influence of internal and external forces.

Offshoring: The relocation of a business process or entire manufacturing facility to a foreign country.

Receiving Organisation: A wholly owned subsidiary or a third party facility in a foreign country where a business process is being relocated to.

This work was previously published in Managing Global Supply Chain Relationships: Operations, Strategies and Practices, edited by Barbara Flynn, Michiya Morita and Jose Machuca, pp. 220-243, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 71
Network Marketing and Supply Chain Management for Effective Operations Management

Raj Selladurai
Indiana University Northwest

ABSTRACT

The use of network marketing by an organization as a distribution strategy in its global supply chain management activities is increasing and becoming very popular. This chapter looks at the "new" network marketing-supply chain paradigm, and analyzes some implications of the strategies for effective operations management. The network marketing paradigm has been greatly enhanced by the growth and popularity of the internet, which provides a powerful impetus to the whole concept of effective global network marketing-supply chain management area. This mainly theoretical study also provides a practical perspective by discussing manufacturing and service companies that are implementing the network marketing-supply chain strategy model in some form. It analyzes the supply chain management activities used by these organizations to achieve unprecedented success in their respective industry. This chapter concludes by developing an effective network marketing-supply chain management model that uses network marketing and supply chain management as key strategies in operations management.

INTRODUCTION

Business organizations across the globe have used a variety of network marketing strategies and integrated them into their operations in many creative ways, to achieve great success. Many businesses have effectively used network marketing strategies that network marketing's reputation, and the popularity of such strategies has grown rapidly in the world over the last several decades. Recently, Warren Buffet's Berkshire Hathaway company got into network marketing when it bought The Pampered Chef, a leading network marketing company. Southern Living at Home, another

DOI: 10.4018/978-1-4666-1945-6.ch071

successful network marketing leader in the home decorating industry, generated more than $100 million a year for AOL Time Warner, its parent company at the time (Woodburn, 2003). Network marketing has extended beyond the U.S. shores, and other countries have begun to take a close look at this "new" phenomenon. This American-style network marketing model has been referred to as the "new" paradigm which is being experimented with and implemented by many businesses and other organizational structures in Japan (Imura, 2004). In China, the world's biggest market, several companies and many local people are getting into network marketing (Rayasam, 2007).

In a recent study, Coyne, Clifford, and Dye (2007) suggested 21 questions that effective companies need to ask, in developing new products. These include specific questions related to buyers and users (knowing the target market and audience well; examining in depth some of the binding constraints (overcoming limitations); exploring unexpected successes (looking at new creative and imaginative uses for the products); moving toward perfection by imagining it (continuous improvement); focusing on and adapting to the external environment ("thinking outside the box" strategies); and taking a closer look at and revisiting products' strengths and weaknesses, opportunities, and threats (mission and vision perspectives). This chapter focuses on a few of the specific questions that leading companies are addressing effectively via specific strategies which have contributed to their success – namely network marketing and supply chain management.

BACKGROUND

Uniqueness and Significance of Network Marketing

Network marketing, also more commonly referred to as direct sales, direct marketing, multilevel marketing and/or other similar concepts, is basically the direct distribution strategy of a product or service, from the company to the consumers, through building effective relationship marketing between distributors and customers. Typically, a network marketing strategy is implemented via a structure or system in which the manufacturing company chooses to distribute its products and services directly to its customers through a network of distributors. This is usually done through a selective group of dynamic and highly motivated people outside the company (distributors, dealers, affiliates, partners, etc.) who market the products and services directly to the consumers and get paid by the manufacturer for doing so. These distributors are also given the opportunity to develop and build their own independent business networks, usually called a downline or multi-level organizational structure, through recruiting, training, and inspiring others to duplicate their efforts. Products are sold mostly from outside of fixed retail locations and usually from the home, primarily through one-on-one personal contact and through product demonstrations. With the increasing popularity of the internet, several companies' dealers and representatives are using the internet to market their products, and are building their business networks by using proven lead generation systems, in collaboration with specialized lead generation companies.

Network marketing, or direct selling overall, may also be referred to as an industry -- a fast growing, successful industry. In 2007 alone, the direct selling industry reported 30.8 billion in total sales in the US, with a workforce of about 15 million involved in some way in this industry (http://www.dsa.org/pubs/numbers/#SALES).
What makes network marketing so popular and unique? The one-to-one personalized, relationship marketing interaction that takes place between distributor and customer is the key to network marketing. Most people would prefer and choose to buy from someone whom they know, like, and trust – network marketing personifies this concept extremely well. People would generally prefer to

buy their product/s from a local distributor – a friend, neighbor, associate, coworker, peer or colleague, fellow member of a church, social club, etc. - rather than from an impersonal corporate office or company sales office. Direct selling or network marketing promotes this personalized touch to selling and buying amongst distributors and customers. Recent studies showed that about 74 percent of Americans have purchased goods or services through the direct sales method. This is more than the number who have purchased through television shopping and on-line computer services, combined. People seem to highly value the products available through direct selling channels and about 45 percent of Americans want to buy from direct sellers (http://www.dsa.org/aboutselling/benefits/). Direct selling is increasingly used as a viable alternative or supplement to traditional marketing. As growth and expansion in the traditional retail, catalog, and online selling channels slows down and often stagnates, many companies are using direct selling to stimulate growth (Duffy, 2005).

Another strong reason for network marketing's growth and popularity is the cost advantage that it provides. Cateora (1990) characterizes channel maintenance costs as the margins, mark-ups, and commissions of channel members in the network. Reducing or eliminating the middlemen in the supply chain so as to allow for greater margins, would be the ultimate goal of manufacturers and traders, and has been from the beginning of trade. The potential to increase profits by eliminating intermediate channels of distribution is a strong motivating factor for the growth of direct marketing. Although network marketers often boast about their channel's efficiency in avoiding wholesalers and conventional retailers, this channel is not a completely low cost channel (Croft & Woodruffe, 1996). However, it still is a profitable source to the manufacturer, as this channel has the ability to command a higher price for its products, compared to similar products which are distributed through traditional retail stores (Eisenberg, 1987).

The advantages that network marketing offers, in terms of capital requirements and cash flow implications, would contribute further to its attractiveness. This channel symbolizes the ideal, where the entire distribution network in the supply chain pays cash with orders and holds inventory at its own expense. For example, Mary Kay Cosmetics encourages its distributors to have a minimum of $600 worth of inventory and this may be increased by a distribution up to around $3000, in order to support her customers. Amway has generated significant cash flows from its operations by requiring its distributors to hold inventory (at least 3 to 8 weeks' stock) and pay upfront for their inventory items (Croft & Woodruffe, 1996).

Network marketing also provides more control and promotes quicker action in the supply chain. Companies can easily communicate with the entire distribution network today through email, online platforms and messages, telephone, and video conferences and can effectively manage the physical distribution of their products through company controlled warehousing. Hence, communication is fast and effective, which enables the quick and efficient introduction of new products, through the distributors in the supply chain, to their customers all over the country and even the world – in a relatively short time and at very lost costs. Network marketing stimulates fast action, compared to traditional retail outlets, in terms of product introduction, sales, after-sales service, handling returns, and prompt delivery of goods and services (Kaikati, 1993).

Network marketing also provides excellent coverage related to maximizing sales volume, market share, and market penetration (Cateora, 1990). Weitzen (1993) stated that network marketing companies stimulate hypergrowth because they can effectively inform, educate, and persuade customers and distribute products quickly and at relatively low cost, especially by using the internet. Further the close relationship building that network marketers focus on and develop with their customers enhances concentrated efforts and activities within particularly social groups, regional, religious, or ethnic community groups. For example, Amway's Malaysian network marketing operations grew faster in the minority Chinese community, compared to other groups (Amway, 1993). EcoQuest International, another leading network marketing company in the health wellness/healthy living industry, had a strong majority Christian group amongst its 80,000+ distributor network.

The unique characteristic of cross-cultural appeal of the network marketing channel is another strong contributing factor for its widespread success (Yarnell, 1994; Weitzen, 1993). It has universal appeal and works in a variety of cultures. Amway's strategies have exhibited remarkable success in Japan where extended family relationships and lifelong close relationships are especially well suited to the direct marketing methods employed by the company (Amway, 1994). Further, a "retail revolution" took place in Japan that enhanced network marketing strategies (Sanghavi, 1990). More Japanese women were employed in the workforce, leading to less shopping hours available, and more of the younger people bought products based on trust, very typical of the network marketing channel. Also, more people were willing to experiment with newer shopping ways such as direct marketing, and a greater number of the aging population was looking for convenience in shopping, perhaps from home.

The entrepreneurial spirit that network marketing encourages is appealing to many who are interested in independent entrepreneurial activities outside traditional business settings (Gabriel, 1993). Many middle-aged Japanese managers, as well as managers from other nationalities in many parts of the world, are looking for independence from the typical employer-employee relationship at work and are venturing out on their own exploring nontraditional self-managed businesses. Network marketing provides a very viable opportunity in this area. When a firm uses a variety of network marketing strategies as tools for distribution and combines these with sustainable supply chain management activities, it is more likely to move toward achieving effective operations management and sustainability.

Significance of Supply Chain Management

Effective supply chain management, or sustainability in supply chain networks, focuses on a sequence of organizations, including facilities, activities, and functions, all integrated toward efficient production and delivery of products and services continually over a long period of time. Sustainability should be measured by not only profits, but also by the impact of the supply chain on ecological and social systems (Jennings & Zandbergen, 2005). An organization that performs well should be focused on doing well on traditional profit and loss measures, as well as on the expanded conceptualization of performance which includes social and environmental dimensions. This is often referred to in the supply chain literature as the triple bottom line (Kleindorfer et. al, 2005).

According to the literature on sustainability and supply chain management, the triple bottom line is affected by three major factors – existing best practices, supply chain strategy, and integration. Several studies have shown a connection between existing best practices in supply chain management, and environmental outcomes and practices. Positive relationships have been found between TQM (Clark, 1999) and JIT (King & Lenox, 2001) and environmental outcomes. Best practices related to sustainable supply chain management include collaboration between supply chain members, cooperation with suppliers, and certifications. Collaborative behaviors between suppliers and customers can enhance environmentally sustainable supply chains (Carter & Carter, 1998; Zhu & Sarkis, 2004). Practices that promote collaborative behaviors include incentives for suppliers to reduce their risk for trying novel collaborative practices required for the sustainability in the supply chain (Goodman, 2000). Firms should share knowledge and educate their suppliers, and these suppliers should educate other suppliers within their supply chain (Rao and Holt, 2005). Supplier certification, which denotes higher levels of quality, is another practice that promotes sustainable supply chain management (Teuscher, Gruninger & Ferdinand, 2005).

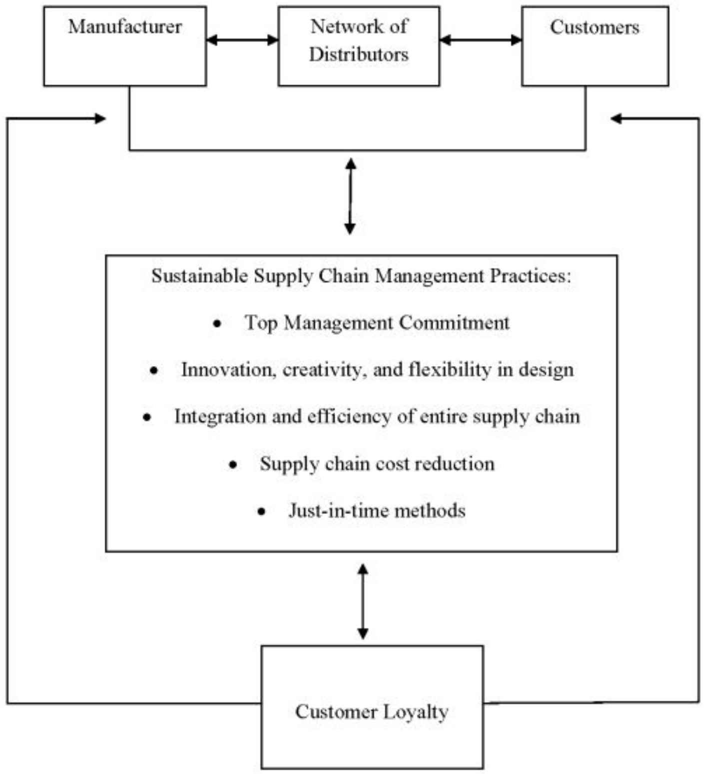
Supply chain strategy has a major impact on supply chain sustainability. Literature studies have described the use and significance of an ecocentric perspective for achieving sustainability -- referring to the organization's ability to adapt well to its social and natural environments (Seuring, 2004). The more flexible and adaptable the organization to its environment, the better its potential for achieving supply chain sustainability. This would allow the organization to include nongovernmental agencies, community groups and agencies, and perhaps even competitors which have traditionally been ignored. Collaboration amongst various members, including nontraditional members in the supply chain, enhances supply chain sustainability (Johnson & Linton, 2000). The use of innovation by a firm as a strategic option enhances sustainability. Sometimes, a firm can be innovative and adopt a service oriented strategy, providing more than the minimum level of service that would be traditionally provided. For example, some chemical suppliers offer chemical management services to other companies in their supply chains (Sharma & Henriques, 2005). They work with the buying firms and provide additional services such as helping design the production facility, facilitating with delivery and handling of materials, and perhaps even helping with waste treatment. Supply chain members are more like partners in a new, mutually beneficial relationship, where all members work together to enhance sustainability for the entire supply chain.

Another factor than enhances sustainability in the supply chain is the integration of sustainability goals, practices, and attitudes into day-to-day activities of supply chain management. Sustainability is everyone's responsibility within the supply chain, starting from top management of every firm. Firms that take a proactive stand toward the environment become more sustainable (Klassen & Whybark, 1999). Further, top management must show a strong commitment, along with taking a proactive stand toward sustainability, which often gets implemented as a written environmental policy (Ramus & Steger, 2000). Other factors that promote sustainability include improved measurement and design processes. Effective supply chain managers design improved measurement processes that show a connection between certain behaviors and desired outcomes, leading to rewards that employees value. This motivates employees toward improved performance, as well as achieving sustainability goals. Redesigning products and processes to eliminate or reduce pollution has shown to be associated with improvements in plant performance (Klassen & Whybark, 1999). It all begins with top management's willingness to provide a strong commitment to achieving sustainability in the supply chain. This may be enhanced by redesigning products and processes geared toward sustainability. Sustainability is made possible only with the cooperation of all employees, who will become involved if motivated properly with appropriate, relevant, and attractive incentives. Therefore, supply chain sustainability can be blended with network marketing by an organization to achieve positive outcomes, specifically customer loyalty, as illustrated in Figure 1.

Effective Network Marketing-Sustainable Supply Chain Integration (ENM-SSCI) Model

Figure 1 illustrates an integrated model, developed and discussed in the following section. This model blends together the strengths of network marketing strategy and supply chain management strategy. The model, referred to as the Effective

Figure 1. Effective Network Marketing-Sustainable Supply Chain Integration Model (ENM-SSCI)



Network Marketing-Sustainable Supply Chain Integration (ENM-SSCI) Model, focuses on the manufacturer using a network marketing strategy, via a network marketing structure or distribution channel, to interact with its customers. The network marketing distribution channel implements certain sustainable supply chain management strategies, which when blended together contribute toward establishing customer loyalty. Positive and strong customer loyalty, in turn, impacts the manufacturer-network marketing channel-customer chain thereby providing a continuous improvement aspect to the model.

The manufacturer makes a conscious decision to specifically choose a network marketing strategy and/or distribution channel because of multiple advantages, including the one-on-one personalized touch between distributor and customer, the cost benefits to the manufacturer in employing this distribution structure, capital requirement and cash flow advantages, improved and quicker action in terms of communication, promotions, new product introduction, improved coverage related to maximizing sales volume, market share, and market penetration, cross-cultural appeal that enables targeting expanded markets; and the entrepreneurial spirit that is encouraged amongst the distributors. All these advantages lead toward motivating stronger performance by the distributors and also enhance manufacturer-distributor relationships. The sustainable supply chain management strategies that are recommended in this model include top management commitment, design innovation, creativity and flexibility, supply chain integration and efficiency, overall supply chain cost reduction, and just-in-time methods for high performance in the supply chain.

To achieve sustainable supply chain management, top management has to be proactive and provide support and commitment to supply chain activities (Klassen & Whybark, 1999). Commitment and a proactive orientation are established through the integration of the economic and non-economic aspects of sustainability. Environmental and social goals should complement and tie in with economic goals and activities, such as revenue generation, profit, marketing, communication and distribution. A proactive orientation is exhibited through organizational attitudes toward sustainability. In a proactive organization, sustainability is part and parcel of the firm's tasks and activities and the mindset throughout the organization and at all times -- 24/7. All decisions made at a proactive and sustainable organization must impact the triple bottom line in some way. In addition, a proactive and committed firm focuses on guiding values or a philosophy related to sustainability that drives the entire organization and its decision making. Responsibility for sustainability must be shared amongst all members of the supply chain – it has to be a shared responsibility for the whole organization and not the responsibility of any one particular person. In this way, organizational sustainability is every organizational member's passion – when people have input in the decision making process for achieving and maintaining sustainability, they are more likely to implement it.

Innovation, creativity, and flexibility in supply chain design also enhance sustainability. These new practices include traceability and transparency (Roth, et. al, 2008). Traceability is an internal information sharing practice to facilitate supply chain performance and minimize risks in the whole chain. For example, information on ingredients or materials used in products in the entire chain, including all items that the supplier bought, should be provided and traced to the source, in order to ensure meeting the specifications and standards for quality and other aspects. Transparency refers to the access/information provided about the flow of money and all related activities, such as profitability, throughout the supply chain. Sustainability is enhanced through information sharing via transparency and traceability in the entire supply chain (Lee & Whang, 2000).

Sustainability is also improved through integration and efficiency of the supply chain by use of strategies such as maintaining supplier continuity and loyalty. As Japanese supply chain management practices have shown, supplier continuity and loyalty are key elements of sustainability (Liker & Choi, 2004). Companies that have continuity in long-term relationships with their suppliers have improved efficiency and integration, as well as the efficiency and integration of the entire supply chain. Sustainable companies treat their suppliers as family members and ensure keeping them successful and in business, as well as helping them to invest, innovate, and grow. Long-term, continuous supplier partnerships imply the use of fewer reliable suppliers, long-term relationships, continuity, loyalty, sharing of information, and cooperation in planning, all of which contribute toward the sustainability of both firms and their suppliers. Further, supplier continuity reduces supplier risk and offsets higher costs, as supply chain members help each other and ensure reliability in the supply chain. Local firms close to the manufacturer are supported and encouraged via long-term contracts and relationships through these supplier continuity strategies, thus providing mutual benefits to the manufacturing firms and their suppliers. Sustainability is enhanced through supplier certification. Supplier certification such as ISO 9000, the most widely used international certification, may provide the status of world class suppliers to firms that obtain it. A major advantage of using certified suppliers is that the buying firms can eliminate all or most of the inspection and testing costs of delivered goods, thus benefitting from lower costs of inspection and testing as well as saving time. Rather than checking and testing the supplies after they arrive, the manufacturing firm can use the items straightaway in their production processes, when they have arrived from a certified supplier. Honda Motors and Toyota use certified suppliers for many components, and both benefit tremendously, especially in terms of lower costs.

Another supply chain practice that promotes sustainability is the implementation of cost reduction strategies and tradeoffs when structuring a supply chain (Lee & Billington, 1992). The lot size-inventory tradeoff, when managed well, can lower the overall cost for both suppliers and the firms. Producing in large batches or ordering large lots provide benefits in terms of quantity discounts and lower setup cost; however, the carrying costs are higher for suppliers which carry these large inventories. An optimal tradeoff here can benefit both suppliers and firms. Similarly, the inventory-transportation cost tradeoff should be balanced. Suppliers prefer to ship large full truckloads to optimize their costs, but this leads to higher inventory carrying costs for customers. Optimal solutions include combining several orders, decreasing truckload capacity, cross-docking, where storage at the warehouse is eliminated, thereby incurring zero storage costs. Several companies including Wal-Mart have used a cross-docking strategy to offset transportation and storage costs. Lead time-transportation cost tradeoffs are enhanced through providing improved forecasting information to suppliers, which helps them in turn improve the timing of their production and orders to their suppliers. The product variety-inventory tradeoff may be optimized through delayed differentiation, which offsets the higher costs associated with product variety. Delayed differentiation minimizes production costs through standardization, rather than product variety and customization, in the production process. Similarly, the cost-customer service tradeoff may be influenced by shipping directly from warehouse to customer, by-passing the retail store. This saves on lead time and reduces storage costs, although it may increase transportation costs. Just-in-time methods enhance sustainability in supply chains. Dell Computers, for example, uses lean manufacturing to maintain sustainability in its supply chain. Dell uses a tiered approach, in terms of its suppliers, which means that, rather than buying from hundreds of suppliers, it buys from a small number of suppliers. It works with few first-tier suppliers which work directly with Dell or which supply major subassemblies. The first-tier suppliers take responsibility for managing their second-tier suppliers which supply components for subassemblies, thereby allowing the final buyer, Dell, to avoid dealing with many suppliers. This type of just-in-time structure enhances speed of delivery, quick and smooth flow of materials, efficiency, zero or little inventory, lower costs, improved quality, and overall increase in productivity, which benefits the entire supply chain.

The effective network marketing-sustainable supply chain management integration model has implications for several factors, including efficiency, profitability, growth and customer loyalty. This chapter focuses especially on customer loyalty -- an essential element for continuous and repeated business, which contributes to overall performance, profits, and growth (Reichheld, 2006). Customer loyalty is greatly enhanced, due to the strong relationships developed between manufacturers, suppliers, distributors, and customers through effective network marketing. This integrated relationship marketing is well illustrated in practice by the Coca Cola Company. The world famous soda company has successfully maintained its major strength for a long time, a strong distribution excellence, which it achieves through providing excellent customer/client satisfaction directly and indirectly to its customers. Strong customer satisfaction is generated by providing one-on-one personalized service to its direct customers and intermediaries – its highly trained network marketing professionals and relationship building experts such as wholesalers and retailers -- who are in constant close touch at the local levels with their final customers and end users. These strategies emphasize Coke's focus on customer service, customer satisfaction, personal stake, and sincere interest in the customer, which customers are looking for in today's business environment. Such a customer-oriented focus and ongoing strong customer service gives Coca Cola the distinctive competency and edge to stay ahead of the competition.

The loyalty factor is also illustrated by Chick-fil-A, the leading fast-food chicken franchise in the U.S. The turnover rate for operators at this chain is five percent a year in an industry which has about a 40 percent turnover rate. Chick-fil-a managers are proud of the way they treat their employees and consider themselves to be very successful, compared to their competition like McDonalds and Burger King, despite working six days a week (Chick-fil-A is closed on Sundays due to religious reasons). Loyalty leaders like Chick-fil-A, The Vanguard Group, and Enterprise Rent-A-Car have all grown at more than double the rate of their competition and enjoy a 15 to 20 percent cost advantage, which further enhances their growth and profit potential (Reichheld, 2006). Thus, the economics of loyalty are tremendously powerful in any industry. It is simple economics though – customer loyalty leads to repeat customers, and repeat customers mean more sales and revenue, which then leads to more profits and growth. What affects customer loyalty? Good customer service drives customers to become loyal repeat customers, and these repeat customers influence their friends and others to become customers too. Network marketing, together with sustainable supply chain management practices, lead to customer loyalty. Reichheld (2006) recommended some ways for building customer loyalty which many successful companies follow. One way includes win-win strategies and solutions – a company has to find processes, strategies, and solutions aimed at mutually beneficial outcomes for itself and its customers. Dell exemplifies this win-win strategy. Dell provides some unique customer services such as bundling software and making it available by shipping directly to the customer's desk. Dell's Premier Pages provide corporate accounts and their personnel a direct access to company-specific online information for delivery related information.

Another approach is to be selective with customers. Loyalty leaders cannot possibly give the world's best value to everyone, as this is almost impossible to do so. Enterprise Rent-A-Car illustrates this strategy. The average rental cost for a car at the airport from Enterprise was about half the price charged by its leading competitors like Avis and Hertz (Reichheld, 2006). Yet, Enterprise is a leader in its market. This is possibly due to its strategy of focusing on special agreements Enterprise has made with selected customers like State Farm Insurance and its clients, who due to prior good experience with Enterprise, will give the company a first try. This facilitates growth for Enterprise through its selective efforts and loyal customers. Another loyalty strategy is making the business simple and flexible in its customer service and transactions. The size of the team structure is a big factor in company flexibility. For Enterprise, Vanguard, and eBay, the ideal team structure is around seven to ten people – many of their competitors are structured larger and are inflexible. Across these successful companies, the relatively small team size make a positive contribution toward customer loyalty.

Loyalty-building companies use sensible and motivating incentives and rewards to enhance loyalty in customers and employees. Unlike some companies that may often try to attract new customers by offering them all kinds of incentives while ignoring their loyal customers (as they are hooked already or trapped), some innovative loyalty leaders like The Vanguard Group do just the opposite. Vanguard rewards its 10 years+ loyal customers by offering them special incentives including cutting their prices by one-third and other better deals that it does not offer to its short term customers.

Another innovative loyalty building strategy focuses on open, direct, communication between customers who help one another share information about the products and/or company and who interact with one another to solve the related problems affecting them. Cisco Systems, eBay, Microsoft and others all offer open communication message boards or similar platforms for their customers to interact, discuss, listen and talk with one another – this approach enhances customer loyalty amongst their customers, especially since this communication is free, current, updated, and available online. Further, companies must not only practice what they preach, but also preach what they practice. They must communicate to their loyal customers what they are loyal to – core values, principles, mission, vision, etc. and run their business according to what they have stated as their guiding principles. The most successful companies, including Chick-Fil-A, Hobby Lobby, Southwest Airlines and Enterprise, all follow this strategy.

Reichheld (2006) pointed out that companies need to focus on the "promoter" type of customers. These are the loyal customers who come back as repeat customers. For McDonalds, 80 percent of its sales revenue is obtained from 20 percent of its loyal repeat customers who eat at their restaurants at least three times per week. The "undecided," the "passive" and the "detractor" types of customers (all of whom do not help the company much) should have as minimal a focus as possible. The bottom line is that company growth is enhanced when a company has more "promoters" than its competition – more long-term repeat customers lead to higher loyalty, thus leading to higher sales, revenue, profits, and growth.

Some examples of progressive companies using variations of network marketing and sustainable supply chain management practices include Continental Power Corporation (CPC) based in Pennsylvania, and EcoQuest International based in Dallas, Texas. CPC, a partial network marketing company with one level of authorized distributors, is an environmentally conscious "green" company. It manufactures and markets commercial, residential, and industrial electrical products that improve the efficiency of electrical system, protect against spikes and surges, and reduce electrical noise, thereby prolonging the life of devices (Personal interview with Allen Johnston, Vice President, Continental Power Corporation, July 15, 2009). CPC has installed one half million systems nationwide and provided reduced consumption of electricity, saving 42,250 tons of coal, 231,500 barrels of oil, 940M cu ft of natural gas, reduced the emission of221,500 pounds of sulphur dioxide, 42,000 pounds of nitrous oxide, and 98,350 tons of carbon dioxide (Allen Johnston, 2009). CPC uses sustainable supply chain practices, such as being environmentally proactive in pursuing environmental, social, and economic goals and a variation of the just-in-time distribution system where large custom product systems are built on demand, usually within two weeks of receiving the order from a commercial client.

EcoQuest International, another environmentally conscious "green" company, produces and markets, through the network marketing model, a variety of environmental and healthy living products, including air, water and laundry purifiers. EcoQuest uses a variation of the network marketing-supply chain management system to distribute its innovative products directly to consumers. Other unique practices include providing rapid and almost instant delivery, excellent customer service, quality products and services, building customer loyalty and strong interpersonal relationships with customers and consumers. EcoQuest seeks environmental, social, and economic goals and focuses on activities that enable it to be a socially responsible company. In 2001, EcoQuest International exhibited corporate social responsibility by donating air purifier units that were used to help in the cleanup operations after 9/11 by the Department of Defense and other US government agencies (www.ecoquestintl.com).

Issues, Controversies, and Problems

The network marketing strategy is sometimes been considered controversial, due to the negative connotations it evokes in some people's minds. The controversial aspects of network marketing may occur due to some random abuse or misuse by some unscrupulous individuals who may unnecessarily hype or exaggerate the benefits of a network marketing strategy. However, this strategy, in itself, is not necessarily bad – like any other marketing strategy, when used properly it can reap adequate or even extraordinary benefits and rewards for networking marketers which implement their strategies effectively. Another problem with a network marketing strategy is that it is not meant to be "one strategy fits all" strategy. It is more suitable for specific consumer products, such as food related products, cosmetics, nutritional items, household appliances and goods, all of which benefit from a personalized touch, to stimulate buying behaviors. Further, such items lend themselves to being easily marketed over the internet, which enhances their distribution strategy in terms of speed of delivery, easier availability of information, products, and related services, and faster promotion of new products and services. Some suppliers have also been criticized for their practices, in terms of excessive profiteering and lack of a cooperative spirit. They often are guilty of minimizing transparency and traceability issues, both of which are essential to sustainability in the supply chain.

When implementing this model, organizations should be careful to ensure that the overall integration of the model is a critical outcome. The model works together, with each component complementing the others in a strong integrative manner. Also, it is a continuous improvement model, in that the entire model's components are constantly interacting and affecting one another toward greater effectiveness and sustainability. To make this implementation more feasible, organizational leadership (often provided by the manufacturing company) is mandatory. Whether it is Amway, EcoQuest International, Tupperware, Pampered Chef or any other network marketing company, the manufacturer/network marketing company should take responsibility for providing supply chain leadership.

Many network marketing companies operate in an international environment. Their network marketing strategies are usually standardized across the countries (with each company establishing and promoting its own unique network marketing strategies) as this is more cost-effective, but supply chain management practices are generally customized to each country's local culture, in order to ensure smooth supply chain integration and management within that county. For example, Amway in Malaysia or Japan follows the standardized network marketing principle of expecting its distributors to hold a certain required amount of inventory, however, that amount varies and is customized to adapt to the local culture. The specific attributes, experience, qualifications, and other selection criteria of the supply chain members vary from country to country and should be customized to fit each country's needs, laws, practices, etc.

A major difficulty is that this model is one-dimensional, in that it can either fully follow a traditional distribution strategy or the network marketing strategy – a blend of the two in the same country would be highly complex and virtually impossible. If a manufacturer wishes to adopt a blend of the two distribution strategies, then two separate divisions should be created to follow separate strategies. Companies that have chosen to go the network marketing route all the way and in every country they operate in believe that this strategy is more beneficial to all parties involved and very rewarding to themselves, their distributors, as well as to their customers.

SOLUTIONS AND RECOMMENDATIONS

To address these issues, a stronger integrative approach is necessary between the network marketing strategy and the supply chain. More cooperative efforts, coordination, and working together are necessary for the benefit of the entire network marketing organization and all of its supply chain members. Constant monitoring and evaluation, to ensure accountability and responsibility of all members of the entire supply chain, are essential to ensure smooth functioning of the whole system and to maintain sustainability within it. Effective leadership must be provided, usually by the manufacturer, to promote sustainability of the supply chain. The larger the organization becomes, the more complex it is to manage it effectively, especially in a cross-cultural and global context. It is highly recommended that such a large manufacturer/network marketing organization adopt some type of enterprise resource program (ERP) software, such as SAP, in order to effectively manage the whole realm of operations across the world. Enterprise resource programs help coordinate and integrate all functions and divisional units of the large organization and help promote improved decision making, communication, information-sharing, productivity, and growth.

FUTURE RESEARCH DIRECTIONS

As this chapter focused on a theoretical model of network marketing-supply chain management, it is suggested that future research studies should focus on empirically testing this model fully or at least major aspects of it. Also, testing this model in different global settings will provide a more comprehensive understanding of the model and its applications in cross-cultural contexts. It would also provide a more logical rationale for generalizing the conclusions and implications of the model discussed in this study.

CONCLUSION

This chapter has looked at the effective network marketing-sustainable supply chain management model and its implications for organizations especially in terms of customer loyalty. Using unique and innovative strategies for network marketing, blended together with sustainable supply chain management practices, an organization can pursue environmental, social, and economic goals which lead to significant outcomes, such as customer loyalty. Positive and strong customer loyalty, in turn, will impact the manufacturer-network marketing channel-customer chain, thereby providing a continuous improvement aspect to the model.

REFERENCES

Amway (1993). *Prospectus for the Issues of 10,000,000 American Depository Shares, Amway Japan Limited*. New York: Morgan Stanley & Co..

Carter, C., & Carter, J. (1998). Interorganizational determinants of environmental purchasing: Initial evidence from the consumer products industry. *Decision Sciences*, *29*(3), 659–695. doi:10.1111/j.1540-5915.1998.tb01358.x

Cateora, P. (1990). *International Marketing* (7th ed.). Homewood, IL: Irwin & Co.

Clark, D. (1999). What drives companies to seek ISO 14000. *Pollution Engineering*, 14–15.

Coyne, K., Clifford, P., & Dye, R. (2007). Breakthrough thinking from inside the box. *Harvard Business Review*, *85*(12), 70–84.

Croft, R., & Woodruffe, H. (1996). Network marketing: The ultimate in international distribution? *Journal of Marketing Management*, *12*, 201–214. doi:10.1080/0267257X.1996.9964409

Duffy, D. (2005). Direct selling as the next channel. *Journal of Consumer Marketing*, *22*(1), 43–46. doi:10.1108/07363760510576545

EcoQuest International. (2009). Retrieved from http://www.ecoquestintl.com.

Eisenberg, R. (1987). The mess called multi-level marketing. *Money*, *16*(6), 136–160.

Gabriel, G. (1993). Land of rising profits. *Success*, *40*(2), 14.

Goodman, A. (2000). Implementing sustainability in service operations at Scandic Hotels. *Interfaces*, *30*(3), 202–214. doi:10.1287/inte.30.3.202.11653

Imura, Y. (2004). Meet the new paradigm: Network marketing in Japan. Can American style marketing work in Japan? *Japan Inc.,* November, 38-44.

Jennings, P., & Zandbergen, P. (2005). Ecologically sustainable organizations: An institutional approach. *Academy of Management Review*, *20*(4), 1015–1052. doi:10.2307/258964

Johnston, T., & Linton, H. (2000). Social networks and the implementation of environmental technology. *IEEE Transactions on Engineering Management*, *47*(4), 465–477. doi:10.1109/17.895341

Kaikati, J. (1993). Don't crack the Japanese distribution system – Just circumvent it. *The Columbia Journal of World Business*, *28*(2), 34–45. doi:10.1016/0022-5428(93)90040-V

King, A., & Lenox, M. (2001). Lean and green? An empirical examination of the relationship between lean production and environmental performance. *Production and Operations Management*, *10*(3), 244–256. doi:10.1111/j.1937-5956.2001.tb00373.x

Klassen, R., & Whybark, D. (1999). Environmental management in operations: The selection of environmental technologies. *Decision Sciences*, *30*(3), 601–631. doi:10.1111/j.1540-5915.1999.tb00900.x

Kliendorfer, P., Singhal, K., & Van Wassenhove, L. (2005). Sustainable operations management. *Production and Operations Management*, *14*(4), 482–492. doi:10.1111/j.1937-5956.2005.tb00235.x

Lee, H., & Billington, C. (1992). Managing supply chain inventory. *Sloan Management Review*, 65–68.

Lee, H., & Whang, S. (2000). Information sharing in a supply chain. *International Journal of Technology Management*, *20*(3/4), 373–387. doi:10.1504/IJTM.2000.002867

Liker, J., & Choi, T. (2004). Building deep supplier relationships. *Harvard Business Review*, *82*(12), 104–113.

Ramus, C., & Steger, U. (2000). The role of supervisory support behaviors and environmental policy in employee 'ecoinitiatives' at leading edge European companies. *Academy of Management Journal*, *43*(4), 605–628. doi:10.2307/1556357

Rao, P., & Holt, D. (2005). Do green supply chains lead to economic performance? *International Journal of Operations & Production Management*, *25*(9), 898–916. doi:10.1108/01443570510613956

Rayasam, R. (2007). Selling, Beijing's way. *U.S. News & World Report*, *142*(8), 52–53.

Reichheld, F. (2006). *The Ultimate Question; Driving Good Profits and True Growth*. Boston: Bain and Company.

Roth, A., Tray, A., Pullman, M., & Gray, J. (2008). Unraveling the food supply chain: Strategic insights from China and 2007 recalls. *Journal of Supply Chain Management*, *44*(1), 22–39. doi:10.1111/j.1745-493X.2008.00043.x

Sanghavi, N. (1990). Non-store retailing in Japan: A huge and potentially lucrative market. *International Journal of Retail and Distribution Management*, *18*(1), 19–23.

Seuring, S. (2004). Integrated chain management and supply chain management comparative analysis and illustrative cases. *Journal of Cleaner Production*, *12*, 1059–1071. doi:10.1016/j.jclepro.2004.02.006

Sharma, S., & Henriques, I. (2005). Stakeholder influences on sustainability practices in the Canadian forest products industry. *Strategic Management Journal*, *26*, 159–180. doi:10.1002/smj.439

Tuescher, P., Gruninger, B., & Ferdinand, F. (2005). Risk management in sustainable supply chain management (SSCM): Lessons learnt from the case of GMO-free soybeans. *Corporate Social Responsibility and Environmental Management*, *13*, 1–10. doi:10.1002/csr.81

Weitzen, H. (1993). Billion-dollar growth. *Success*, *40*(3), 14.

Woodburn, M. (2003). Tapping the network marketing model. *Direct Marketing*, October, 1-4.

Yarnell, M. (1994, March). Worldwide boom. *Success*, *41*(2), 23.

Zhu, Q., & Sarkis, J. (2004). Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. *Journal of Operations Management*, *22*, 265–289. doi:10.1016/j.jom.2004.01.005

KEY TERMS AND DEFINITIONS

Network Marketing: Also more commonly referred to as direct sales, direct marketing or multilevel marketing and/or other similar concepts, is the direct distribution strategy of a product or service from the company to the consumers through effective relationship marketing between distributors and customers.

Sustainable Supply Chain Management: Effective supply chain management or sustainability in supply chain networks focuses on a sequence of organizations, including facilities, activities, and functions, all integrated toward efficient production and delivery of products and services continually over a long period of time. Sustainability should be measured by not only profits but also by the impact of the supply chain on ecological and social systems.

This work was previously published in Managing Global Supply Chain Relationships: Operations, Strategies and Practices, edited by Barbara Flynn, Michiya Morita and Jose Machuca, pp. 322-335, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 72
Knowledge Management in SMEs:
A Mixture of Innovation, Marketing and ICT: Analysis of Two Case Studies

Saïda Habhab-Rave
ISTEC, Paris, France

ABSTRACT

Global economy is transforming the sources of the competitive advantages of firms, especially for firms embedded in local manufacturing systems. Based on the theoretical contributions to knowledge management and industrial districts, this paper describes alternatives firm's strategies and upgrading options by exploring the relationships among innovation, marketing and network technologies. Starting from the analysis of the global competitiveness report and the European Innovation Scoreboard, this paper focuses on the case of firms specializing in "furniture and textile" industries (fashion, mode, home products) to outline a framework explaining the new competitive opportunities for SMEs. Through a qualitative analysis, this paper presents two case studies of French firms that promote successful strategies based on a coherent mix of R&D based innovation, experienced marketing and design, by leveraging on ICT.

INTRODUCTION

Global economy is transforming the sources of firms' competitive advantages and especially for firms embedded in local manufacturing systems. As in the case of France, small and medium enterprises (SMEs) localized in industrial districts and specializing in low or medium-tech industries have built their success on productive flexibility, quality certification and incremental innovation. Literature on industrial districts has provided evidence of the sources of competitiveness of local systems (Porter, 1990). As opposed to large multinational corporations, district SMEs emphasize an alternative model of economic organization (Porter, 1998), in which external economies

DOI: 10.4018/978-1-4666-1945-6.ch072

support distributed production processes within the local networks of firms. From this perspective, on the one hand, scholars focused on the advantages offered by proximity in terms of technology spillovers and economic externalities (i.e., Krugman, 1991) (collective goods). On the other hand, studies on the knowledge economy (Arora et al., 1998) consider industrial districts as knowledge management systems, where the local context is able to sustain and facilitate creation, exploration and exploitation of knowledge, rooted into social practices (Nonaka and Takeuchi, 1995).

In the actual environment, SMEs are now facing competitive forces that impact on the sustainability of their strategies in the next years. First, manufacturing internationalization pushes firms operating in local supply chains to extend their networks beyond local boundaries to catch the opportunities of global value chains (Gereffi et al., 2005). While, on the one hand, a growing part of local productive activities may be transferred internationally with cost advantages, on the other hand, those paths may reduce a small firm's control over economic processes with negative influence on learning-by-doing innovation.

A second major challenge refers to the development and management of sales networks on a global basis, in a framework of stronger connections with the market. As many scholars have outlined, the interaction between customers and the firm through sales networks, as well as the web, is crucial in order to understand the market and anticipate demand trends. More important, building relationships with active customers (lead users and communities of customers) is part of a firm's innovation strategy, to obtain profitable knowledge for product and brand management (i.e. Sawhney, Prandelli, 2000). From this perspective, SMEs have to improve their competencies in interaction with customers at the international level, overcoming local, social and cultural boundaries as well as their traditional manufacturing approach. Such strategic options require more sophisticated marketing competencies, which are not usually available within SMEs operating in local productive systems.

Thirdly, the evolution of information and communication technologies (ICT) contributes to the debate about the transformation of the district firm model and the advantages of local embeddedness. Global supply chains and international commercial outlets ask the firm to increase control on processes at the organizational level and within the firm's extended value system. From this perspective, network technologies can strengthen information sharing, process transparency and interaction among players in the value system (final customers included). Large multinational companies were able to fill the gap with the flexible SME model in the 1990s, thanks to network technologies. These tools supported distance cooperative work, also increasing process monitoring, knowledge management and communication within a renovated firm model (Scott Morton, 1991). In the present scenario, SMEs are asked to update their strategies benefiting from network technologies. SMEs have to overcome the local environment as the prime source of innovation –local tacit knowledge mainly manufacturing-oriented and informally managed- by developing new capabilities to manage extended networks including research centers, designers, and customers.

Based on the theoretical contributions to knowledge management and industrial districts, this paper describes alternative firm's strategies and upgrading options by exploring the relationships among innovation, marketing and networks technologies. The paper focuses on the case of firms specializing in "furniture" and "clothes and shoes" industries (fashion, mode clothes, and equipments, home products) to outline a framework explaining new competitive opportunities for SMEs. Our hypothesis is that the learning-by doing innovation model that has characterized district firms in the past is no longer sufficient to sustain their competitive advantage. The RD based innovation, efficiently adopted in large corporations, can offer new strategic options to face international competition. However, it cannot be implemented easily in all district SMEs. Moreover, innovation cannot be limited to scientific knowledge management, but can benefit also from customer input and experience related to technical features as well as associations and symbols the product incorporates (Krippendorf, Butter, 1984). From this perspective, the capabilities of SMEs to manage networks of relationships and to translate customers' needs into products may open new competitive opportunities, under the condition of a well-defined ICT strategy.

In the first section, this paper analyzes the district SMEs' model and its impact on French competitiveness, based on the contributions and approaches to innovation of the Global Competitiveness Report and the European Innovation Scoreboard. The second section focuses on the drivers of competitive advantage and strategies of firms in terms of science-driven and market-driven innovation, also considering the role of ICT. Through a qualitative analysis, in the third section, this paper discusses two case studies of French firms that promote successful strategies based on a coherent mix of R&D based innovation, experienced marketing and design, by leveraging on ICT.

1. SMEs' COMPETITIVENESS IN THE INTERNATIONAL SCENARIO

Despite scholars' interests in the emerging countries economic model based on competitive local systems of SMEs, international analysis stresses the marginal role of France in the global arena as regards SME's capabilities to manage codified innovation. The Global Competitiveness Report of the World Economic Forum put France out of top 10, 16th in the international ranking. This study emphasizes the dynamics of growth and competitive factors of countries (with a focus on technology innovation, economic systems and institutional framework) through a comparative approach and identifies the competitive potentials of firms localized in each country. As opposed to success during the '80s and '90s, French economic system, and specifically SMEs specializing in "furniture" and "clothes and shoes" industries (home products, fashion, equipment…), seem to lack competitiveness, due to low investments in R&D and patents. Even in the European Union Framework, the tool used to evaluate competitiveness and performances of nations and regions – the European Innovation Scoreboard- describes a quite negative picture of French firms, based on a few indicators on firm's expenditure on R&D, the numbers of patents registered, and investments in advanced services.

As stressed by analysts, the prevalence of small and medium firms in the economic system is the principal reason for French weakness in managing innovation successfully. According to the data of European researchers, the French SMEs are characterized by learning-by-doing innovation. Thus, SMEs are not able to translate new knowledge into patents and codified outputs. Moreover, SMEs do not approach (formal) innovation with strategic intent and, hence, do not invest a relevant amount of resources in R&D, training and new technologies. Despite this negative picture, those studies mention a few French SMEs' strengths related to organizational innovation and strategic control on technical activities such as product design.

From our perspective, the explanation of such contradictory results can be explained by considering a broader approach to innovation, which does not cover only R&D based activities. Instead, innovation can also be linked with the development of intangible features of the product and customer experience as the main drivers of value creation. From this standpoint, there are many different ways through which innovation can be deployed: the value created through innovation and its impact on competitiveness is rooted in the variety of forms and processes of the innovation each firm is able to design in its own original way. Following this approach, recently, the European

Union has upgraded its framework of analysis by creating the Innovation Diversity Index, which is a measure oriented to capture the alternative forms of innovation characterizing countries and regions. Such an index is influenced not only by innovative firms that invest in R&D and patents, but also firms that have positive performances based on organizational innovation and innovation in marketing and design.

From this point of view, the competitive advantage of France becomes clearer. Despite their specialization in low or medium-tech industries, French SMEs and their leadership in the area of technological innovation (16th in the innovation pillar) are important attributes that have helped to boost the country's growth potential with regards to innovation management processes that develop and transform informal knowledge into value for the market. France's labour market flexibility continues to be ranked very low (131st) because of the rigidity of wage determination, high nonwage labour costs and the strict rules on firing and hiring. Another area of concern is macroeconomic stability (65th): the government budget deficit and the related public-sector debt ratio remain large, and the national savings rate, while growing, still remains low by international standards.

In this scenario, of near formalized procedures that lead to innovation –typically used in large corporations- one should also evaluate, on the one hand, the openness of the innovation cycle (innovation inputs beyond scientific knowledge and R&D) and, on the other hand, the results of innovation (outputs) and its use. Based on the Innovation Diversity Index of the EU, French SMEs show strong ability in the management of networks and collaboration. Traditional innovation drivers (R&D, skilled labor force and lifelong training) are weak in SMEs (ranked 21 out of 25). Instead, small firms are stronger in new knowledge generation and implementation.

According to the categories developed by the EU, French firms are classified as "modifier" in their innovation strategy because they capture and transform external knowledge into products through informal processes. Such approach is perceived either negatively, as it is not codified (and represents incremental innovation) or positively, as SMEs are flexible in knowledge management. Firms can reinvent products and processes in many original ways thanks to their reactivity to market inputs and demand and by developing differentiation strategies. This capability is supported by specific professional practices focused on product specialization available at the territorial level. We explain those results by referring to the economic district model, where small businesses belonging to local networks of production organize knowledge management through distributed innovation systems, instead of a large organization.

During the fordism paradigm, the large firm model has been considered the best way and scientific knowledge (and R&D) was the main driver of innovation. In the open innovation paradigm, distributed networks sustain innovation (Chesbrough, 2003) and customers can contribute with their knowledge (Von Hippel, 2005). Moreover, customers are available to pay for products that offer not only new features (technological innovation), but also which offer them an experience and the tangible value linked to associations with sensemaking supported by brand strategy, design and social participation (Prahalad, Ramaswamy, 2003). From this perspective, innovation cannot be limited to technological innovation, but should also include aesthetic and intangible elements created through marketing strategy (communication) (Bettiol, Micelli, 2005). According to this perspective, French firms may improve their position in the international competitive arena because of their specific capacity to face innovation.

2. STRATEGIES, KNOWLEDGE MANAGEMENT AND ICT

In low or medium-tech industries such as textile or furniture, the competition is increasing and requires firms to choose either cost leadership in the mass market or niche differentiation, while positioning in the middle-market is becoming more and more unsustainable (Silverstein, Fiske, 2003). As opposed to high-tech industries, in which the role of patents and collaboration with research institutions is crucial for product innovation, in the mentioned industries innovation cannot usually be perceived as patent-driven. Instead, innovation is linked to creativity, a firm's ability to manage variety (innovation as organizational capability), and mix inputs coming from the market, designers and marketing (Schmitt, Simonson, 1997). From this perspective, an evaluation of a firm's innovation performance and its strategy should not be limited only to R&D activities and its outputs. Rather, from our perspective, in the open-innovation paradigm (Chesborugh, 2003) it should also consider the extension and characteristics of the networks that sustain a firm's innovation (as inputs of knowledge) as well as innovation outcomes. Marketing scholars emphasize the role of the intangible as part of the innovation process and a result of the value offered to customers. Products are not sold only because of their new features and functionalities, but also, and often, due to the meaning they transmit through their shapes (design) and the experience they give to customers (Pine and Gilmore, 1999).

Studies on innovation process have stressed the role of codified knowledge in knowledge management cycles, while the analysis of social dynamics (Brown and Duguid, 2000) has outlined the situated learning system and the relevance of experience as a driver to develop and share complex knowledge. According to this point of view, SMEs operating in local manufacturing systems benefit from physical proximity to customers, suppliers and relevant communities of practices embedded into local contexts. However, the global competitive scenario forces SMEs to upgrade and develop new strategies where innovation processes are sustainable on an international level. In a complex and global market, where leading customers are far from the firm and there are numbers of potential knowledge sources for a firm's innovation (Tapscott & Williams, 2007), the local economic and social system is inadequate to offer SMEs all the relevant and useful knowledge to compete. On the one hand, modularity and codification can guarantee a more open and extended circulation and use of knowledge, across contexts. On the other hand, the more complex knowledge to manage, the higher the difficulties in codification and the need for promoting more sophisticated sharing strategies based on "pragmatic collaboration" (Helper, MacDuffie, Sabel, 2000) (people-to-people by face-to-face interaction or web-based).

Based on this distinction and the literature contributions on the topic, we can represent the sources of firms' competitive advantage (Grant, 1996; Kogut & Zander, 1996) by comparing the different role of knowledge developed by firms and the alternative strategies of knowledge management adopted. We identified alternative models (Table 1). On the one side, we can identify firms that compete by leveraging on R&D and scientific knowledge. Codification allows firms to enter into global networks of innovation and exchange knowledge on a broad scale with universities and research centers (regional innovation systems, Asheim Coenen, 2006). Local dynamics are supported by international connections, through which the firm is able to explore opportunities and exploit knowledge. On the other side, competitive advantage is based on customer relationship management built on experience. The firm is interested in selecting lead users and involving customer communities into the innovation processes, aiming at their sharing relevant knowledge (Von Hippel, 2005). It is a form of entrepreneurial innovation, with a strong role of marketing, as the firm's organization and processes are oriented to the market and to interact with external players (customers and lead users) to co-develop the product and the meaning related to it (Muniz & O'Guinn, 2001). Our hypothesis

Table 1. Competitive advantage and firm strategy

Technology	Organization	Marketing
R&D (patents)	Mix of different types of knowledge	Brand strategy design
Human Capital	Management of extended manufacturing and Commercial networks	communication and interaction with customers (lead Users, community)
Technological Innovation	Distributed innovation	
Science-based Innovation	Non technical – Marketing based innovation Entrepreneurial innovation	
←---→		
Focus on research	*Focus on organizational Processes*	*Focus on communication*

is that in the complex competitive scenario, firms may develop sustainable competitive advantage by mixing the strengths of the opposite models, where patents and R&D based innovation may be enhanced through marketing based innovation and vice versa.

- **Technology & R&D:** R&D within the traditional corporate value chain has been regarded as a function closely related to product development or manufacturing. However, there are indications that this is changing through the introduction of knowledge management and the fact that R&D is more often performed within networks across traditional corporate boundaries. Griffin and Hauser (1992a, 1992b) focus their research on the interaction of marketing and the R&D interface. Additionally, customers are rising as potential innovators, which should be included into the R&D process as early as possible. Based on the above discussion we suggest in this paper that a more market- and customer-oriented approach to R&D could be obtained by applying a radically different view of R&D activities. We propose a relationship marketing approach based on knowledge management.
- **Organization:** Attempts to define KM mirror and reflect the intangible, fragmented and multifaceted nature of knowledge itself. KM has thus been defined in a whole host of ways that vary in scope and focus. In terms of scope, the term has been used broadly to refer to "the capacity (or processes) within an organization to maintain or improve organizational performance based on experience and knowledge". In terms of focus, definitions emphasize, variously: organizational processes and routines; performance improvement outcomes; processes for networking and collaboration; practices for harnessing, storing and distributing expertise; specific tools and methodologies such as data-mining and storage systems. These definitions suggest a variety of practices and organizational processes "a skillful blend of people, business processes and IT". However, research and practice in KM has been dominated by a focus on using Information Technologies (Its) to store, search and transfer knowledge within and across organizations. The logic behind this technocratic approach to managing knowledge is that by implementing various kinds of IT (databases, intranets) coupled with relevant search engines, knowledge can be captured and transferred from place to place. The assumption here is that if knowledge is transferred via technology, it can be used for innovation in other parts of the organization and so means that the chances of needless-

ly what has already been done elsewhere will be minimized. The notion that knowledge can be extracted from where it lies, codified and moved, using Its, en masse from one place to another has been challenged for failing to adequately address the problems of managing tacit knowledge. Tacit knowledge by definition is, at best, difficult or, at worst, impossible to articulate, values and experiences that give tacit knowledge its meaning. Moreover, it is precisely because tacit knowledge is tacit that makes it difficult to for other organizations to imitate or import and therefore makes it an important organizational resource for securing competitive advantage. But the cognitive model of KM adopts a partial view of knowledge, assuming that knowledge lies with individuals and largely ignoring the socially constructed and socially mediated nature of knowledge. So, organizational knowledge is thus both widely distributed and embedded in collective systems of meaning and action. The aim is to discover the best practice in KM. with the development of the field of KM there has been a massive outpouring of articles and books dealing with these issues from a prescriptive standpoint. The problem is that many of these best practice prescription assume a direct, functional relationship between knowledge, marketing and innovation, despite evidence to suggest that this is over-simplistic.

- **Marketing:** The importance of marketing for business success is not new. Highly respected scholars like Peter Drucker and Theodore Levitt published their classic writings on these subjects as early as 1974 and 1975 respectively. Drucker (1974) argued that the only purpose for a business is to create a customer and that is achieved through marketing and innovation. Levitt (1975) argued that marketing is largely ignored because top management is wholly transfixed by the profit promises of technological R&D. Levitt also argued that product-orientedness in high technology works well where firms are pushed into new frontiers where they did not necessarily have to find markets but to fill them. Based on the above discussion we suggest in this paper that a more market-and customer-oriented approach to innovation could be obtained by applying different view of innovation activities.

In such competitive scenario, the analysis of a firms' strategy about innovation management cannot be developed without the study of their approach to information and communication technologies (ICT). On the one hand, these technologies support information management at a distance, by stressing the advantages of efficiency. On the other hand, thanks to multimedia tools, ICT allow the development of a virtual, interactive environment, where participants live the experience and are involved in social interaction on line. This environment offers opportunities related to knowledge creation and sharing, even in the case of complex knowledge (i.e. product innovation).

It is not our aim to describe the debate on the impacts of ICT on knowledge management in detail. We would outline the SMEs' approach to ICT investment and its influence on innovation. The international reorganization of manufacturing activities, as well as sales networks, push firms to adopt technological solutions that sustain coordination of activities in extended networks and organizations (Scott Morton, 1991). Moreover, the transformation in the consumption models described above asks firms to interact with customers in order to exploit the linkages with lead users and communities for innovation purposes. In both the strategic options of a science-based, competitive advantage (i.e. patents) and value-driven by "customer intimacy" and sensemaking (Treacy & Wiersema, 1997), network technologies become key factors in supporting competitiveness. In the open-innovation paradigm, ICT is in fact the valuable infrastructure for knowledge management aims, where knowledge is spread across contexts, organizations, and people (employees, customers).

Computers-mediated communication offers tremendous advantages of tracking and tracing dialogues and interactive relationships, as well as content development and sharing (digitalization, multimedia solutions, social software)(Von Hippel, 2005), even in complex situations. Hence, technologies can help firms overcome barriers and leverage the networks of connections characterizing the on-line environment (in primis among customers). Traditionally, ICT found primary application in large corporations, to solve coordination problems and support knowledge gathering and retrieval efficiently and effectively (Sproull, Kiesler, 1991). The role of technological infrastructure as a necessary condition for knowledge management did not match with the SME's competitive model. Especially within local systems, small firms have developed knowledge management mechanisms rooted in the social sphere of their contexts of embeddedness (Becattini & Rullani, 1996). Knowledge processes are usually not codified in formal procedures, but lie in the intensive communication and personal linkages within the organization, as well as outside the working domain, in the social fabric of places.

As shown in studies on ICT adoption in district firms (Chiarvesio et al., 2004), as opposed to large corporations, small and medium firms localized in local manufacturing systems have focused their attention on commodity-based technologies, such as email and web sites. Those technologies can be considered ready-to-use tools, which can be implemented in the organizational structure with low financial investments, as well as limited organizational changes. In industrial districts, SMEs'strategies in ICT investments have been characterized by:

- Selectivity in the technological solutions chosen;
- Incremental innovation processes based on learning-by-doing paths;
- A bottom-up process (no "master mind" at the local level).

During the new economy many scholars and analysts stress the potentialities of e-commerce for SMEs in terms of market enlargement and efficiency. Instead, research on ICT adoption by French district SMEs show rates of e-commerce, while the web is exploited as an interactive marketing tool. Firms do not consider the e-commerce solutions available adequate to manage "French" products for transactional purposes. Rather, firms stress the importance of web-based communication: the web becomes a medium to gather customers' feedback on products and support brand strategies.

More advanced technologies such as ERP (Enterprise Resource Planning) or groupware, tailored to large firms, are less diffused in small organizations. However, those solutions are considered crucial tools to increase process transparency and the control on distributed networks at the international level. In this perspective, the more extended the firm's value chain, the higher the need for upgrading the SMEs' strategy, where ICT sustains the firm's management beyond the local system. From our perspective, all the technological solutions available can be included in the framework of the knowledge management system, not limited to the organizational boundaries, but involving the players operating upstream (suppliers, designers, etc.) and downstream (sales agents, customers) in the product innovation as well as marketing activities.

3. COMPETITIVENESS IN FRENCH FIRMS

In this theoretical framework, we considered the strategies of firms specializing in low and medium-tech industries to explore the connection between R&D based and marketing-driven innovation processes, and the role of ICT in supporting those activities at the local and global level.

In order to explore the strategies of French firms in the scenario described above, we carried out a qualitative study on district SMEs to analyze knowledge management processes and firms' innovation approach (Yin, 1984). Based on a first selection of firms specializing in "agriculture and textile" sectors and located in north east of France, we interviewed entrepreneurs and the managers of R&D, design and information system departments. Interviews focused on a firm's history and strategy, organizational structure and innovation management models and ICT adopted. The two cases discussed in the paper are summarized in Table 2.

Case 1: "World Champions" in Mixing Design and Patents

One of the most famous companies of the sport system, firm X started in 1973 producing clothes and shoes, followed by shoes and other products for child and younger people. During the '80s the company internationalized its business, thanks to soccer shoes and international partnerships with France and foreign stars. Moreover, firm X is among the first district firms that invested in internationalization of productive activities beyond the local manufacturing system. As a leading firm in Parisian district, firm X invests in innovation to support its competitiveness by coupling R&D-based activities (scientific research on new materials, design, etc.) and the involvement of lead users. One of its latest products is, in fact, a mode of jeans, designed in collaboration with stars. Meanwhile, in a framework of global production and commercialization of products, network technologies have been considered key elements in the management of extended supply and sales networks with the district as the core.

In the new millennium, the attention for the investment in product quality has been increased through an explicit strategy that emphasizes the role of design and innovation as drivers of competitiveness. In the global competitive scenario, firm X is oriented to reinforce its international presence. In this perspective, cost reduction as a key goal to face competition has to be coupled with continuous product innovation. The development of original ideas –jeans – is the starting point in X's strategy. The management of internal knowledge is relevant both in terms of R&D and design – more than 20 patents have been registered or are in the process of registration. The development of research relationships with French and international universities stresses X's interests in exploring knowledge paths beyond the local district networks to sustain the company internationally.

As an open network firm, firm X has invested in network technologies systematically, by gathering different technology solutions – from their web site, to e-commerce, ERP, groupware and supply-chain management applications. The technological infrastructure sustains information flows and communication between the company and its international networks of partners and

Table 2. Case studies

Company	Innovation strategy	Main ICT investments
Case 1: Firm X 230 employees; Sport system: shoes and apparel	Mix of R&D (patent, relation with universities) and design	Explicit management processes
Case 2: Firm Y 40 employees; furniture	Collaboration with external international designers Patents	Internet to support online distributed product design processes

markets, in a strategic and codified knowledge management approach. In fact, firm X is interested in acquiring and sharing informal knowledge available within the organization through ICT (digital archives, database for intellectual property rights management).

Case 2: From the District to International Design Network

Firm Y is a small firm specialized in the production of high quality furniture and wooden complementary house products. It was founded in 1989 and is located in Paris in France, in the furniture district. Firm Y has developed its strategy by focusing on product differentiation through design. Since 1998, firm Y has been obtaining economic success and growth thanks to the international recognition of a few of the firm's products – the Parisian, a famous French design award promoted by the French design association- designed by one of Y's founders. From these awards, this small firm started relevant collaboration with international designers. Due to investments in developing personal relationships and connections with designers all over the world (Japan, Europe, USA), the firm was able to increase the product range and international sales (60% of the turnover is export-based).

Y's strategy is oriented to exploiting internal strong competencies in wood transformation and production of "natural wood" furniture. The manufacturing process is organized in small-scale stocks, with particular emphasis on product customization as regards to the material used and finishing activities. Specifically, a mix of hand-made and technological innovation processes characterizes Y's made-to-order production. As opposed to the typical district approach in which local suppliers are key players in the firm's innovation processes, Y has developed innovation mainly internally, through R&D activities and patenting, and is able to increase the technological features of the products as well as their design characteristics. In the global competitive scenario, Y's approach to innovation is double: on the one hand, the focus is on design and aesthetical components of products as drivers of economic success; on the other hand, this small firm invests also in codified knowledge to protect their ideas against competitors (1 to 3 years in the average time of the product innovation cycle). Y does not invest in market research. Instead, the firm exploits international designers and entrepreneur's knowledge about customers and future trends, as an emerging process.

The entrepreneur is confident about the strategic role of ICT to sustain the firm's competitive advantage. Network technologies are key tools to support creativity processes, while the web infrastructure allows Y to interact with its commercial networks. Specifically, the firm's exploitation of multimedia applications and broadband opens new opportunities in product design and development at the international scale. In fact, the product "Riddled" – obtained through collaboration with the famous Steven Hollen's design studio based in new York and produced in 39 plus 30 items – has been made possible thanks to on line communication and document sharing at a distance between France and the USA. At the same time, Y has also created an open and distributed digital archive concerning all the documents and digital contents about products and innovation processes to use them for marketing and knowledge management purposes.

4. CONCLUSION

The two case studies are characterized by successful strategies based on a mix of R&D-driven innovation and marketing, where firms developed strong relationships with customers. Innovation processes blend codified knowledge and tacit knowledge based on specific practices related to consumption or professional profiles (exploitation as well as exploration in knowledge management, J.March, 1991). The firms interviewed are able to couple scientific innovation with product innovation based on design, the creation of experience and focus on communication. The local context in which these firms are embedded is important, but it is not the only source of knowledge in order to build their competitive advantage. On the one hand, these firms are interested in creating new connections with foreign research centers to promote projects for product, technology or material innovation. On the other hand, they develop relevant linkages with the loci of consumption and with key players for creativity, to nurture the innovation process interactively.

The local context offers competencies in the manufacturing domain and sustains the culture of the product. However, competitive SMEs are able to create and manage extended networks by operating in global value chains and approaching innovation through the entrepreneurial innovation model identified by the European Union. To be sustainable those strategies require information and communication technologies, where ERP systems support advanced process management and increase interoperability, while web-based solutions for communication and product (document) management are also implemented in supply chain and commercial sales networks.

Even if our study is still preliminary in its term, the case studies offer a few managerial implications in the way the innovation process is outlined as an open process. First, firms should understand the types of relationships characterizing the players involved in the innovation dynamics, in order to develop consistent mechanisms of management (codification vs. interaction). Second, there are interesting opportunities in combining different kinds and sources of knowledge, which have to be identified and coordinated. Today, firms are asked to develop capabilities in accessing external knowledge (exploration) through people-moving and electronic connections. In addition to this flexibility and openness they also have to pursue strategies and use tools (ICT) coherent with the relationships developed.

REFERENCES

Arora, A., Gambardella, A., & Rullani, E. (1997). Division of labour and the locus of inventive activity. *The Journal of Management and Governance*, *1*, 123–140. doi:10.1023/A:1009993430964

Asheim, B. T., & Coenen, L. (2006). Contextualizing regional innovation systems in a globalizing learning economy: on knowledge basis and institutional frameworks. *The Journal of Technology Transfer*, *31*, 163–173. doi:10.1007/s10961-005-5028-0

Brown, J. S., & Duguid, P. (2000). *The social Life of Information*. Cambridge: Harvard Business School Press.

Chesbrough, H. W. (2003). *Open Innovation*. Cambridge: Harvard Business School Press.

Davenport, T. H., & Prusak, L. (1998). *Working Knowledge. How organizations manage what they know*. Boston: Harvard Business School Press.

Drucker (1974). *Management Tasks, Responsibilities, Practices*. London: William Heinemann LdT.

Global Competitiveness Report. (n.d.). Retrieved from www.weforum.org/pdt/GCR08/GCR08.pdf

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*, 109–122.

Griffin & Hausser. (1992a). Patterns of communication among marketing, engineering and manufacturing – A comparison between two new product teams. *Management Science*, *38*(3), 360–373. doi:10.1287/mnsc.38.3.360

Griffin & Hausser. (1992b). *The Marketing and R&D Interface*. Massachusetts Institute of Technology Marketing Group.

Helper, S., MacDuffie, J. P., & Sabel, C. M. (2000). Pragmatic collaboration: advancing knowledge while controlling opportunism. *Industrial and Corporate Change*, *9*(3), 443–488. doi:10.1093/icc/9.3.443

Kogut, B., & Zander, U. (1996). What firms do? Coordination, Identity, and Learning. *Organization Science*, *7*(5), 502–518. doi:10.1287/orsc.7.5.502

Krippendorf, K., & Butter, R. (1984). Product semantics: exploring the symbolic qualities of form in innovation. *The Journal of the Industrial Designers Society of America*, *3*, 4–9.

Levitt (1975). Marketing Myopia. *Harvard Business Review*, September-October, 12.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–86. doi:10.1287/orsc.2.1.71

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14–37. doi:10.1287/orsc.5.1.14

Pine, B. J., & Gilmore, J. (1999). *The experience economy*. Boston: Harvard Business school Press.

Porter, M. E. (1990). *Competitive advantage*. New York: The Free Press.

Prahalad, C. K., & Ramaswamy, V. (2003). The new frontier of experience innovation. *MIT Sloan Management Review*, *44*, 12–18.

Sawhney, M., & Prandelli, E. (2000). Communities of creation: managing distributed innovation in turbulent markets. *California Management Review*, *42*, 24–54.

Schmitt, B., & Simonson, A. (1997). *Marketing aesthetics. The strategic management of brands, identity and management*. New York: The Free Press.

Scott Morton, M. S. (Ed.). (1991). *The Corporation of the 1990s. Information technology and organizational transformation*. New York: Oxford University Press.

Silverstein, J. M., & Fiske, N. (2003). *Trading up. The new American Luxury*. New York: Portfolio.

Sproull, L., & Kiesler, S. (1991). *Connections. New ways of working in the networked organization*. Cambridge: MIT Press.

Tapscott, D., & Williams, A. D. (2007). *Wikinomics. How mass collaboration changes everything*. New York: Penguin Books.

Treacy, M., & Wiersema, F. (1997). *The discipline of market leaders: choose your customers, narrow your focus, dominate your market*. New York: Perseus Books Group.

Von Hippel, E. (2005). *Democratizing Innovation*. Boston: MIT Press.

Yin, R. K. (1994). *Case study research: design and methods*. Thousand Oaks, CA: Sage.

This work was previously published in Knowledge Management for Process, Organizational and Marketing Innovation: Tools and Methods, edited by Emma O'Brien, Seamus Clifford and Mark Southern, pp. 183-194, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 73
Developments in Modern Operations Management and Cellular Manufacturing

Vladimír Modrák
Technical University of Kosice, Slovakia (Slovak Republic)

Pavol Semančo
Technical University of Kosice, Slovakia (Slovak Republic)

ABSTRACT

Operations management as a knowledge domain appears to be gaining position as a respected and dynamic academic discipline that is undergoing constant development. Therefore, from time to time it is sensible to monitor and analyze its developments by summarizing new features into comprehensive ideas. To support this necessity, the major publications/citations in this field and their evolving research utility over the decades are identified in this chapter. Because the goal of this book is to present the advancements in the area of operations management research, especially of advanced topics related to the layout design for cellular manufacturing, the second part of this chapter is focused on developments in cellular manufacturing approaches and methods by mapping literature sources during the last decade. Finally, the relationships between concept or/and tools in both areas that are empirically considered as consequences or coincidences are identified.

INTRODUCTION

Although the overviews of detailed historical developments in each cognition domain are useful, this survey will discuss modern eras of operations management and cellular manufacturing in a successive order.

DOI: 10.4018/978-1-4666-1945-6.ch073

Operations management (often called production management) may be defined in different ways depending upon one's attitude or point of view. Since this discipline is a field of management, it focuses on carefully managing processes to produce and distribute products faster, better and more cheaply than competitors. Operations management (OM) practically concerns all the operations within the organization and the objectives of its activities focus on the efficiency and effectiveness of processes. The modern history of production and operations management was initiated in the 1950s by the extensive development of operations research tools such as waiting line theories, decision theories, mathematical programming, scheduling techniques and other theories. However, the material covered in higher education was quite fragmented without the umbrella of what is called production and operations management (POM). Subsequently, the first publications 'Analysis of Production Management' by Bowman and Fetter (1957) and 'Modern Production Management' by Elwood Buffa (1961) represented an important transition from industrial engineering to operations management. Operations management finally appears to be gaining a position as a respected academic discipline. Thus, this may be a good time to update the evolution of the field. To achieve this goal, the major publications/citations in this field and their evolving research utility over the decades will also be identified in this chapter. Subsequently, opportunities and challenges of a modern operations management that managers were facing during the last decade will be examined.

Because the goal of this book is to present the advancements in the area of operations management, especially advance topics related to the layout design for manufacturing environments, the second part of this chapter focuses on developments in cellular manufacturing approaches and methods. A large body of literature has attracted a number of researchers to present different reports on the state of the art at different points in time. Several researchers have reviewed the literature and categorized the different methods. Our intention in this chapter is to analyze production-oriented cell formation methods based on the review mapping literature sources from 2000 to 2010.

Finally, in this chapter, we will note the relationships between concept or/and tools in both areas that are empirically considered as consequences or coincidences.

OPERATIONS MANAGEMENT IN THE CONTEMPORARY ERA

The process of building operations management theory and the definition of its scope or area has been treated by a number of authors. As mentioned above, the modern era of POM is closely connected with the history of industrial engineering (IE). The development of the IE discipline has been greatly influenced by the impact of operations research (Turner et al. 1993). Operations research (OR) was originally aimed at solving difficult war-related problems through the use of mathematics and other scientific branches. The diffusion of new mathematical models, statistics and algorithms to aid decision-making had a dramatic impact on industrial engineering development. Major industrial companies established operations research groups to help solve their problems. In the 1960s, expectations from OR were extremely high, and as was commented by Luss and Rosenwein (1997), "over the years it often appeared that the mathematics of OR became the goal rather the means to support solving real problems." This caused OR groups in companies to be transferred to traditional organization units within companies. As a reaction to this disappointment Corbert and Van Wassenhove (1993) classified OR specialists into three classes: theoreticians, management consultants, who focus on using the available methods to solve practical problems, and the "in-between" specialists called operations engineers, who adapt and enhance methods and approaches in order to solve practical problems. The term "operations engineers" was formulated due to the lack of a better term and accordingly the group could also be referred to as operations managers and the field conducting applied research to help solve practical problems could be named production and operations management. In further developmental stages of OR, the term POM was consolidated and presented as concepts, methods and approaches related directly to productive systems and enhancing their management. Based on such a derivation of the mentioned disciplines it is obvious that IE, OR and OM have commonalities and similarities in their definitions. However, it is also important to specify the main differences among them. OM is a field of management, OR is a branch of applied mathematics (AM) and IE is an engineering discipline (Chase and Aquilano, 1989). In addition, according to Anderson (2002), OM and OR differ substantially, since "OM is managerially and activity oriented while OR is mainly technique and mathematically oriented involving modeling a situation or a problem and finding an optimal solution for it." Figure 1 illustrates a conceptual map of the compared disciplines with two axes: Symbolic/Real and Analytic/Synthetic.

As can be seen from the figure above, OM and OR follow two complementary routes that create a win-win scenario. Fuller and Martinec (2005), based on their analysis of the parallels between OM and OR, mentioned that both disciplines can be considered as "innovations" of the twentieth century.

Development Features of Operations Management

To discuss the important features of operations management, the reasonable action is to define and explain how the term can be understood from the viewpoint of the book's theme. For this purpose the following definition can be adopted: Operations management is concerned with the ways of achieving the most effective and efficient use of an organization's resources to produce goods and services needed by customers. It goes without saying that there are many other definitions that are more or less similar to the above definition. Although Chase and Aquilano (1989) precisely documented the historical development of OM starting with its real roots given by Taylor (1911), operations management is as old as industry itself (Bicheno and Elliot, 1997) and was articulated in the context of industrial production only after the 1960s (Baber, 1996; Landes, 1998). Because Chase and Aquilano in their above-mentioned book mapped

Figure 1. Conceptual map of relations between OM, IE, AM and OR

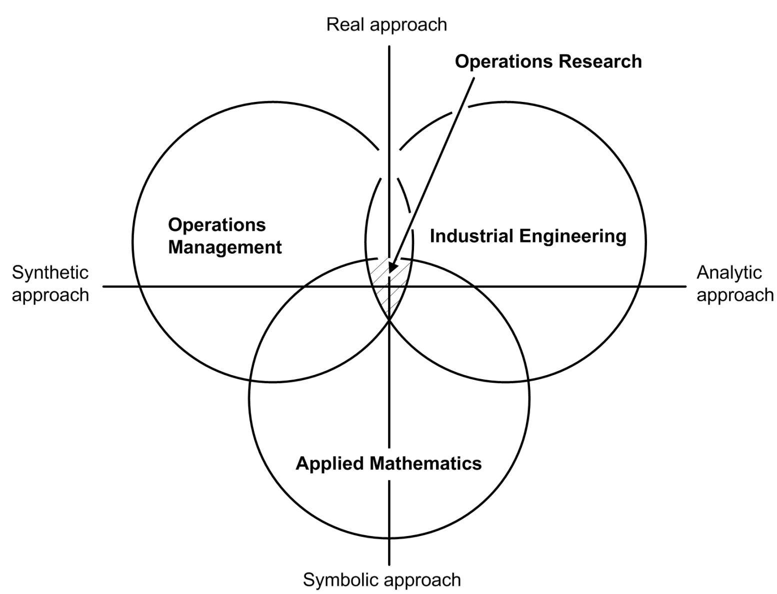


Table 1. Historical summary of the history of modern OM (adopted from Chase and Aquilano, 1989)

Decade	Concept or Tool	Originator or Developer
1950s	Extensive development of OR tools of simulation, Queuing theory, Decision theory, Project scheduling techniques of PERT and CPM	David Georg Kendall (UK) Erich Leo Lehmann (USA)
1960s	Mathematical programming in industrial applications, Extensions of linear programming	Tjalling Koopmans (Netherlands) George Bernard Dantzig (USA)
1970s	Software packages for: Shop scheduling problems, Layout design, Forecast methods, Material requirement planning (MRP)	Joseph Orlicky (USA) Oliver Wight (USA) George W. Plossl (USA)
1980s	JIT, TQC and TQM, Factory automation, CIM, FMS, CAD/CAM Manufacture resource planning (MRP II)	Taiichi Ohno (Japan) W. Edwards Deming (USA) Armand V. Faigenbau (USA) Joseph M. Juran (USA) Mikell. P. Groover (USA)
1990s	The principles process innovations and business process reengineering, Logistics and Supply Chain Management, Optimized Production Technology (OPT), Theory of Constraints,	Thomas H. Davenport, (USA) Michael Martin Hammer (USA) James A. Champy (USA) Richard J. Schonberger (USA) Martin Christopher (UK) Eliyahu M. Goldratt (Israel) Eliyahu M. Goldratt (Israel)
2000s	Management of technology change, Disruptive Innovations and Organizational Change Operations strategy,	Clayton M. Christensen (USA) Wick Skinner (USA) Slack Nigel (UK)

the history of the field of OM between circa 1910 and 1990, the following Table 1 gives an updated view of the historical development of operations management from the beginning of its modern era. The aim of this overview is to trace the key concepts and tools by decades from 1950 to 2010.

We are aware that the above specification of the latest developments in OM may not fully encompass all the decisive development directions, since operations management is a fair-sized and diversified field.

Opportunities and Challenges of Modern Operations Management

During the latest decennium operation managers had to react to unforeseen situations more frequently than they had needed to before. In this context they were facing new challenges of market globalization, information and communication technology advances as well as opportunities for organizational improvements and efficiencies.

One of the important challenges for the development of modern OM was the emergence of so-called global growth companies (GGCs). According to Jones (2005), during the mid-nineteenth century, thousands of European companies were formed exclusively to operate internationally with no prior domestic business. Those companies became momentous actors of economic globalization in international business. The emergence and influence of a new breed of high-growth global companies paying special attention to China and India was discussed by Kiggundu and Ji (2008). Concentrating on China and India as representatives of emerging economies explains the fact that these two countries were the best represented at the Dalian meeting in 2007 organized by the

Centre for Global Growth Companies. By then, "firms in emerging economies and developing countries tend to have weaker systems of corporate governance than those in developed economies." In this connection, findings from differences between emerging economies and developed economies provide excellent opportunities for the study of corporate governance among global growth companies.

The aim of specifying general decisive and substantial challenges that managers have faced during the last 10 years comprises a substantial task, as it depends on different aspects. In an attempt to complete this task, in the Table 2 we depict some topical challenges that are related to the latest concepts and tools shown in Table 1.

In the continuing text, some of the main features of accented concepts and tools assigned to the last decade (shown in the table 1) will be illustrated with the aim of proving their topicality.

Management of Technology Change

According to Thomas and B. Grabot (2006), two main factors have dramatically changed the industrial context in the manufacturing area: specialization and technological changes that have recently occurred in the information technology area. Attention to that fact along with a large diffusion of innovations in industries during the twentieth century most likely evoked the emergence of the new managerial discipline of management of technology (MoT). The term itself was first introduced at the European Management Forum held in Davos in 1981. There are several definitions of MoT, which differ in their understanding of the very object of technology management in the sense of what needs to be managed. Drejer (2002), in this context, commented that "the discipline of MoT is characterized by a vast number of contributions emerging in a divergent manner rather than a convergent one." A succint definition of MoT has been formulated, for example, by Bueno et al. (1997), according to whom it is "the

Table 2. Selected challenges of modern operations management

Challenge	Description
Global Competition	Global market is increasingly complex and constantly changing. Products are traded internationally and components are sourced internationally. It requires a greater degree of international and cross-cultural communications, collaborations, and cooperation than at any time before. All companies have to think in global terms as regional companies are rapidly becoming a thing of past. (Steers and Nardon 2006).
Developments in strategic management approaches	Hambrick and Fredericson (2001) in their paper have talk about their uncertainty of whether that most organizations do actually have a strategy. According to them a meaningful strategy might consist of five elements, providing answers to following questions: Where will we be active? How will we get there? How will we win in the market-place? What will be our speed and sequence of moves? How will we obtain our returns? In reality, most strategic plans emphasize one or two of the elements without giving any consideration to the others.
Supply Chain Standardization and Integration	During the last decade has been proved the slogan that, much competition occurs between supply chains, not just between individual firms. This is due to the fact that company can't act as isolated entity, but as a part of supply chain integrated system.
Complex external environments	It is of crucial importance to understand how external environment impacts on organization. Therefore, companies are quite interested in knowing about macro environment situation representing the information on trends for demography, market geography, technologies energy demand growth, labor productivity growth, etc.. The environment in a global economy and its interactions with organizations is not only a complex phenomenon, but it is constantly changing in nature. Accordingly, any aspects of the environment can't be study as deterministic entities. By Kazmi (2008), "the organization and the environment are, in reality, more unpredictable, uncertain and non-linear". Therefore, for their study the complexity theory including chaos theory and their applications are applicable.

combination of competences allowing technological capabilities aiding the achievement of business objectives to be promoted and controlled." Although definitions of MoT are specific to a concrete target platform, the main object of interest in this chapter is its relevant context to business activities. One of the important roles of MoT is to promote innovation. It is especially topical for organizations that face a serious problem when technological changes are necessary in response to market signals. The internal conditions for implementing advanced technology for routine production are not always adequate for achieving this aim. Draft (2010), in this context, saw a problem with the organization of work. He argued that this problem can be solved only through innovative-oriented organization, which is typically associated with change and is considered the best for adapting to a changing environment. Therefore, programmes for the development of employees' creativity have become an important element of a cohesive corporate strategy.

Disruptive Innovations and Organizational Change

Presently, distinctions between disruptive technologies versus sustaining technologies are frequently discussed. According to the findings of Bowers and Christensen (1995), disruptive changes in technology had a significant impact on industries and many leading companies failed when they were confronted with them. Paradoxically, these failed firms were well-managed companies that invested aggressively in new technologies, carefully studied market requirements and opportunities and sharpened their competitive edges. Christensen (2002) proposed five principles of disruptive technologies in order to find a way to understand and harness this phenomenon. In his fourth principle he focuses on an organization's capabilities and disabilities, stating that "to succeed consistently, good managers have to be skilled not only just in choosing, training, and motivating the right people for the right job, but in choosing, building and preparing the right organization for the job as well." So, it is axiomatic that the phenomena of disruptive innovations and management of technology change are mutually reinforcing.

Operations Strategy

Admittedly, the operations or manufacturing strategy is considered as an inherent part of the long-term corporate strategy. Chase et al. (2004) offered with his sketch of a short history of operations strategy a broader insight into current operations strategy research and determined its role in contributing operations management functions to a firm's ability to achieve its competitive advantage in the marketplace. Since a firm's strategies are often changing and developing, it implies making sensible decisions that affect the business performances directly. In that context Swink and Way (1995) saw the position of manufacturing strategy as "the decisions and plans affecting resources and policies directly related to the sourcing, production and delivery of tangible products." Slack and Lewis's (2002) view of operations strategy is that it is not only a single decision, but the total pattern of the decisions that include the extent and ability of its capacity; delivery of products and services; approach to developing process technology, etc. The importance of operations strategy follows on from the fact that the long-term success of manufacturing firms depends on their ability to vary their operations quickly enough to fill the changing requirements of customers. The key factor that makes the operations function faster is called the manufacturing vision. For this reason, all principal world-class manufacturers have explicitly formulated a strategic manufacturing vision. Practically, it means that all the decisions related to system design, planning, control and supervision made by shop-floor managers are consistent with a corporate vision. On the other hand, world-class manufacturing ambition is not the only issue that matters. Therefore, it is not always optimal to adopt the most offensive manufacturing concepts that are inherent in world-class manufacturers. Accordingly, investing in improving marketing activities, product design or manufacturing operations can be as effective.

DEVELOPMENTS IN CELLULAR MANUFACTURING APPROACHES AND METHODS

Cellular manufacturing (CM), considered as an application of GT philosophy and its principle that focuses on the identification of similar parts to the benefit of a particular production, offers promising alternative solutions for manufacturing systems. CM can essentially be comprehended as a strategy that divides machines and parts into small groups or cells, where each cell can produce a family of parts completely. The manufacturing cells are basically composed of the heterogeneous machines to produce particular families' parts, which are allocated to these cells. For this purpose, various approaches and methods have been developed. Moreover, the CM approach benefits both the job and mass production. The main enhancement of cellular manufacturing implementation incorporates reductions in set-up time, throughput time and material handling and improved quality management.

One of the basic problems that has to be solved before implementing CM is the cell formation problem (CFP). The objective of the CF is to establish the family of parts and the group of machines for subsequent processes. The process of cell formation differs with respect to whether manufacturing cells have been created by rearranging existing facilities on the shop floor or whether new facilities are acquired for the cells. During the decades, a significant amount of research papers have been devoted to this problem. In this regard, several attempts at the classification of CF methods have been introduced. The classification of CF approaches was introduced for instance by Offodile et al. (1994) and Irani et al. (1999) In order to generalize previous classification frameworks, Table 3 shows a basic categorization of CF approaches that is in accord with the already-introduced contributions. The cell formation methods in Table 3 are accordingly divided into three basic groups. Each group has its own direction for part or machine identification.

In the last four decades, CMS research has mainly focused on production-oriented approaches. Therefore, the comprehensive reviews and taxonomy of studies that are devoted to CFP have been presented in previous research works. The following authors participated with their research studies in arranging all these CF methods into groups based on criteria like the minimization of inter-cell moves, machine utilization and others. Wemmerlöv and Hyer (1986) categorized more than 70 papers into 4 representation groups. Subsequently, Selim et al. (1998) reviewed the lit-

Table 3. General classification of cell formation methods

Category	Brief description
Visual inspection methods	Visual inspection methods or eyeballing rely on the visual identification of the particular part families and machine groups.
Part classification and coding methods	PCA-based methods are oriented to design or shape feature. They attempt to group identical or similar design and manufacturing attributes into families. Therefore, they are ideal for reduction of product variety.
Production-oriented methods	The aim of these methods is to apply principles of line production to other types of production than mass production, even when the output is small and there is a large diversity of product. The PFA-based methods seek the optimal solution of cell formation in regard to objective and constraints.

erature (from 1963 to 1998) aimed at the cell formation problem, which is considered a fundamental issue in the CM environment. A comprehensive mathematical formulation of the CF problem has also been presented. The classification of reviewed papers based on multi-criteria cell design was employed by Mansouri et al. (2000), who applied the number of criteria as a measure for classification. They presented a review regarding the multi-criteria objective decision models that take into consideration the manufacturing cell formation problem.

Another view of the proposed taxonomy framework was introduced by Papaioannou and Wilson (2009). They also provided a review and comparison of 52 CF methods.

Production-Oriented Methods for CFP

Following the literature, the main scope within manufacturing cell formation methods focuses on production-orientated methods. In this section we present a modified classification framework for production-orientated CF methods and approaches based on the previous ones as shown in Figure 2.

In association with the proposed classification a brief review of the production-oriented CF methods in the following particular subsections is presented.

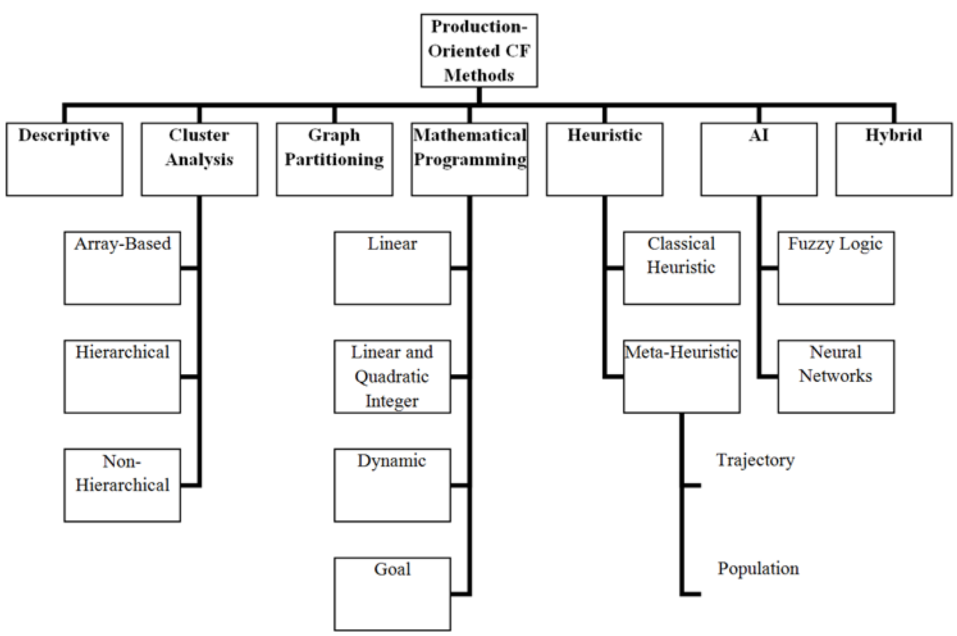
Descriptive Methods

One of the earliest descriptive approaches was developed by Burbidge, who began with the wave of CM systems. The proposed method by Burbidge (1971) is referred to as production flow analysis (PFA). The aim of PFA is to analyze the information from route cards to form cells. A manual method of Burbidge's PFA (1977) for CFP solutions is named nuclear synthesis.

Cluster Analysis Methods

Another group of approaches is presented as cluster analysis (CA). The objective of these methods is to group objects or entities or their attributes into clusters. Diverse techniques are applied in order

Figure 2. Classification of production-oriented methods for CF problem



to create clusters. Those techniques can be further classified as array-based clustering methods, hierarchical clustering methods and non-hierarchical clustering methods within their own subsection. The first one rearranges the order of rows and columns to find a block-diagonal structure of a machine-part incidence matrix (MPIM). They are also known as array-based methods. McCormcik et al. (1972) are counted among the first researchers to have developed the cluster analysis method for the CM environment. They introduced the bond energy analysis (BEA) method. The other well-known CA-based methods are rank order clustering (ROC) by King (1980), direct clustering analysis (DCA) by Chan and Milner (1982) and modified rank order clustering (MODROC) by Chandrasekharan and Rajagopalan (1986). The block-diagonal structure (BDS) uses the minimization of the number of exceptional elements (EE) as its objective. EE represent inter-cell moves, which means in practice that they imply undesirable movement of parts to machines between the individual cells. An exceptional element basically means a bottleneck machine allocated to a cell while it is required in the other cells simultaneously, or a part in a family that requires the capabilities of machines allocated to other cells. The traditional MPIM is also represented as the binary (zero-one) matrix. The rows of the MPIM are machines and columns stand for parts. The entries in the matrix are '0's and '1's, which indicate whether a part needs to be a machine for production or not. The mathematical formulation to create a binary machine-part incidence matrix is defined as follows.

$$a_{ij} = \begin{cases} 1 & \textit{if part } j \textit{ visits machine } i, \\ 0 & \textit{otherwise.} \end{cases} \tag{1}$$

where 'i' is the machine index (i = 1, 2, 3,…, M), 'j' the part index (j = 1, 2, 3,…, P), M stands for the number of machines, and P the number of parts.

Figure 3 a presents the initial machine-part incidence matrix with a size of 9 machines and 17 parts that is formed by Equation (1). Figure 3 b shows the block-diagonal structure that includes one exceptional element.

The hierarchical methods are aimed at the separation of the MPIM data in several stages. In the first stage the MPIM data are grouped into a few broad cells. Subsequently the broad cells are partitioned into smaller groups until terminal groups are generated. The most frequent hierarchical methods applied to cell formation have been

Figure 3. Illustration of (a) binary MPIM with size of 9x17 and (b) block-diagonal structure with one EE

Machine \ Part	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0
2	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1
3	0	0	0	0	0	1	1	0	0	1	0	1	1	0	0	0	0
4	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1	1	0
5	0	0	0	0	0	1	1	0	0	1	0	1	1	0	0	0	0
6	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1
7	1	0	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0
8	0	0	0	0	0	1	1	0	0	1	0	1	1	0	0	0	0
9	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1

a)

Machine \ Part	1	4	9	15	16	20	2	3	5	8	11	14	17	6	7	10	12	13
1	**0**	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	0	0	0	0	0	0	0	**1**	0	0	0	0
7	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0
6	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0
9	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

b)

the three linkage methods. Single linkage (SL) was used by McAuley (1972), average linkage (AL) was used by Seifoddini and Wolfe (1986) and complete linkage (CL) was used by Mosier (1989). A representation of the hierarchical methods can be made by inverted tree structures also known as dendograms. The last of the cluster analysis methods, non-hierarchical methods, are iterative methods that need an initial partition of the data set. One of the well-known methods is ZODIAC, developed by Chandrasekharan and Rajagopalan (1987).

Graph Partitioning Methods

Graph partitioning methods consider machines or parts as nodes that are connected by arcs that represent the production flow between the machines. Graph methods enhance other methods like cluster analysis methods. Rajagopalan and Batra (1975) proposed the method that combines the use of a similarity coefficient and graph theory to solve the cell formation problem.

Mathematical Programming

Since the 1980s, a large number of research papers have been published in the field of mathematical programming with the aim of solving cell formation problems. Kusiak (1987), with his integer mathematical programming approach, was among the first of the authors to apply these methods to CFP. The formulation of mathematical programming (MP) can be employed to model CMS problems in a number of circumstances concerning a wide range of manufacturing data. The objective of MP is regularly maximization of the total number of part similarities in each cell, or minimization of inter-cell material handling costs. Most of the MP-oriented research papers are introduced and discussed by Selim et al. (1998). MP can be classified into four further groups with regard to their type of formulation: linear programming, linear and quadratic integer programming, dynamic programming and goal programming. Boctor (1991) dealt with the mathematical programming method. He proposed a linear formulation of the machine-part cell formation problem. Mathematical programming approaches belong to very time-consuming methods, which is why researchers have turned their attention to heuristic methods with their implementation in CMS.

Heuristic Methods

The heuristic methods are very fast in contrast to mathematical programming methods or others. It is generally known that the heuristic methods do not guarantee to find the optimal solution. However, if they are properly implemented and tuned up, the solution found will represent the optimum in most cases. They reach an optimal or pseudo-optimal solution in a reasonable amount of time. Heuristic methods start from a feasible solution then generate other random solutions, evaluate them and improve the effectiveness or goodness of the solution as time progresses. The presented classification framework in Figure 1 considers the further division of heuristics into classical heuristic methods and meta-heuristic methods, which incorporate evolutionary-based methods and population-based methods. There are numerous different heuristic approaches that are summarized in published review studies. Some of them are mentioned further.

Artificial Intelligence

Another significant group of methods aimed at CMS is introduced as artificial intelligence (AI) that is inspired by nature itself. Fuzzy logic and neural networks are the main approaches of this group. Kaparthi and Suresh (1992) proposed an application of neural networks to solve the cell formation problems. AI can be used to find patterns in manufacturing data in the CM environment. In most cases artificial intelligence approaches represent a robust and adaptive system. During the learning process they can perform a structure modification based on the information that flows through the network. Yang and Yang (2008) proposed a modified ART1 AI-based method to group data into machine-part cells. Guerrero et al. (2002) introduced the self-organizing neural network (SONN) approach, which solves the CF problem using a two-phase strategy. The first phase is dedicated to part-families formation and the second one assigns the machines to each part-family. Other contributors who have dealt with the artificial intelligence methods are shown in Table 2.

Hybrid Methods

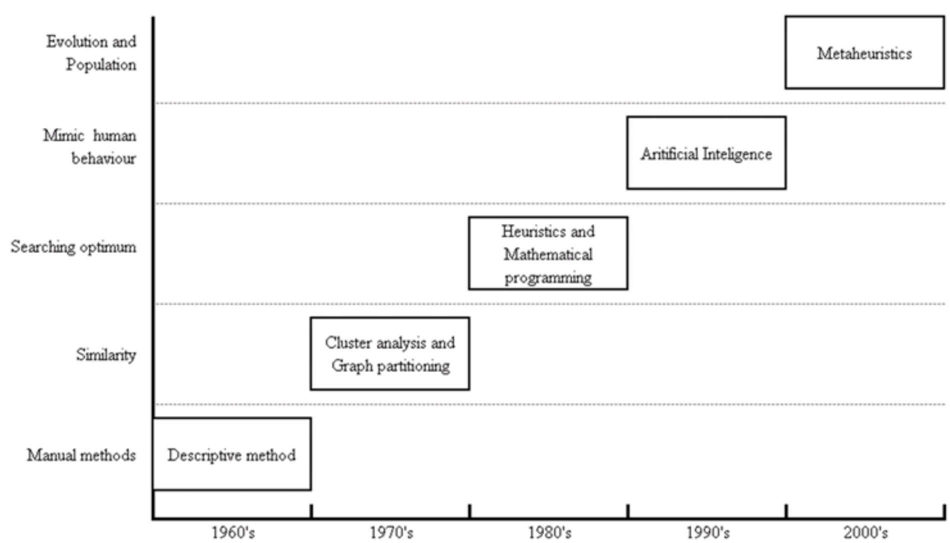
The last group, frequently referred to as hybrid methods, solves the CF problem by combinations of two different methods. In this case they are based on combinations of cluster analysis methods, mathematical programming methods, heuristics and AI approaches. The hybrids take advantage of both methods by applying them to the cell formation problems they can solve efficiently. Based on the literature, in the last decades the hybrid methods have become very popular for the cell formation process in CMS. Caux (2000) proposed an approach that combines both the simulated annealing (SA) and branch-and-bound (BB) algorithms. The proposed approach (SABB) can simultaneously solve CF problems consisting of grouping machines into manufacturing cells and selecting one process plan for each part. Other hybrid methods are incorporated in Table 3.

Review of Modern Methodologies for the Cell Formation Problem

Based on the literature we can introduce an overall view of the evolution of production-oriented methods in CMS. This overall view corresponds with the previously presented classification of CF methods. It is divided into decades, starting with the 1960s, as shown in Figure 4.

Based on the Figure 4 we can propose another classification of CF methods according to progressivity as follows: classical optimization CF methods and modern CF methods. The first category includes descriptive methods, cluster analysis methods, graph partitioning methods, mathematical programming methods and some of the heuristic methods. The second category is

Figure 4. Overall view of CF methods evolution



formed by meta-heuristics, hybrid methods and AI approaches.

Because there are a number of existing studies mapping the time period from 1960 to 2000, our review presented in Table 4 focuses on the last decade. For the purpose of this review a classification based on descriptive approaches, cluster analysis approaches, graph partitioning approaches, mathematical programming approaches, heuristics, artificial intelligence and hybrid methodologies has been applied to categorize recent works. In addition, the methods are reviewed by key elements, such as the maximum size of the problems solved and the performance measures used.

From the survey depicted in the graph (see Figure 5) it is possible to identify the frequency of production-oriented cell formation methods.

The graph shows that the most frequent category of production-oriented cell formation methods is heuristics, including meta-heuristics (category # 5). Heuristics have been employed in CFP due to their promptness and acceptable solutions. Besides the heuristics approach there is another category that has attracted attention over the last few years. This category includes hybrid-based CF methods, which have increased significantly among the researchers in combinatorial optimization. Hybrids exploit the best from both methods combined in one technique. This provides more efficient behavior and higher flexibility. Based on the findings depicted in Figure 5, hybrids along with AI-based CF methods (categories # 7 and # 6) create the second level of importance. The next level of importance includes cluster analysis and mathematical programming methods (categories # 2 and # 4).

Even though the mathematical programming and the heuristic method for the CF problem were firstly introduced in about the same decade, the use and further development of mathematical programming methods in CFP were affected by computational limitations for large-scale problems and they were also time consuming.

The heuristic category has been chosen to compare the usage of particular CF methods due to the fact that it is the most frequent category of all. Based on Figure 6, the most utilized CF methods of the meta-heuristic methods in decreasing order are GA, ACO and TS. The cumulative percentage of the methods used is about 67%. The rest of the methods do not exceed the 10% bound, including the classical heuristic-based CF methods along with the simulated annealing method, scatter search method, particle swarm optimization (PSO) method and water flow-like algorithm (WFA). The order of particular methods in Figure 5 is obviously influenced by the chronology of their development, which has been provided in Table 4. From the mentioned review of the modern cell formation approaches in the last decades it is evident that the scatter search (SS) and water flow-like algorithms (WFA) have just recently received attention for CME.

CONCLUSION

The presented parallel survey on modern operations management and cell formation approaches independently maps the major development features in both POM and CM areas. As shown by Figure 1, CM research also represents a subset of POM. Then, the relationships between concept and/or tools in both areas can be empirically considered as consequences or coincidences. The causal relationship depends on the chronological order and assumes that X precedes Y otherwise Y would not occur. Accordingly, the surveys are based on taxonomies (classes or categories of items) and the chronological timing of the development changes.

Based on the comparison of the two areas it is possible to imply that mathematical programming is a mutual technique for POM and CM, even though this technique was dominantly applied in the given areas in different time periods. It shows

Table 4. Review of the modern cell formation approaches

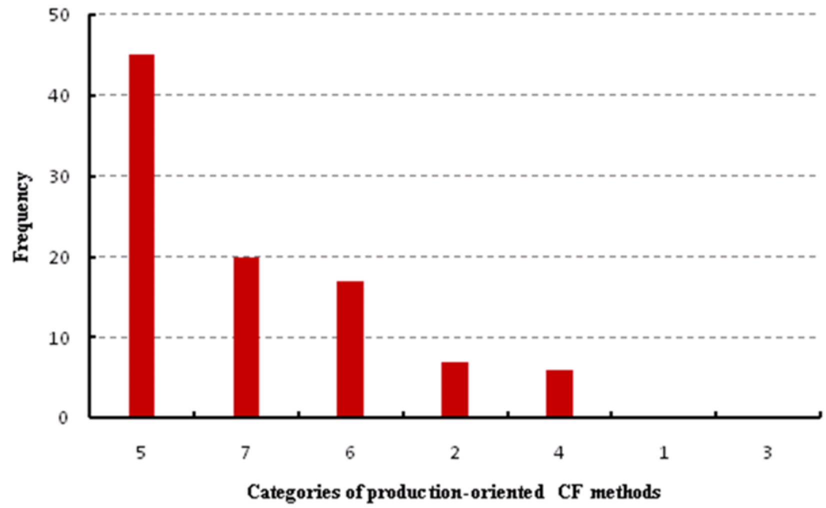
No.	Source		CF method	Categories							Performance			
	Author(s)	Year		1	2	3	4	5	6	7	A	B	C	D
1	Caux et al.	2000	Hy	·	•	□	□	•	□	•	(20x20)	·	✗	
2	Mak et al.	2000	GA	·	·	□	□	•	□	·	(40x100)	·	✗	
3	Lozano et al.	2001	Hy	·	•	□	□	·	•	•	(50x100)	•	✓	
4	Onwubolu and Mutingi	2001	GA	·	□	□	□	•	□	·	(20x45)	•	✓	4
5	Ravichandran and Rao	2001	Hy	·	•	□	□	·	•	•	(9x9)	·	✓	
6	Guerrero et al.	2002	NN	·	□	□	□	□	•	·	(40x24)	·	✓	1
7	Lozano et al.	2002	TS	·	·	□	·	•		·	(50x150)	·	✓	
8	Mukattash et al.	2002	H	·	·	□	□	•	□	·	(13x13)	·	✓	
9	Soleymanpour et al.	2002	Hy	·	·	·	•	·	•	•	(40x100)	•	✓	3
10	Logendran and Karim	2003	Hy	·	·	□	•	•	□	•	(15x30)	·	✗	
11	Park and Suresh	2003	Hy	·	·	□	□	·	•	•	(25x40)	·	✓	
12	Spiliopoulos and Sofi-anopoulou	2003	TS	·	·	□	□	•	□	·	(30x30)	•	✗	
13	Cao and Chen	2004	TS	·	·	·	·	•		·	(8x15)	•	✗	
14	Goncalves and Resende	2004	Hy	·	·	·	·	•	·	•	(40x100)	•	✓	6
15	Kim et al.	2004	H	·	·	□	□	•	□	·	(40x40)	·	✗	
16	Solimanpur et al.	2004	GA	·	·	□	□	•	□	·	(15x30)	•	✗	
17	Won and Lee	2004	Pm	·	·	·	•	·	·	·	(50x150)	·	✗	
18	Wu et al.	2004	TS	·	·	□	□	•	□	·	(18x5)	·	✗	
19	Albadawi et al.	2005	Hy	·	•	·	·	•	·	•	(16x43)	•	✓	4
20	Islier	2005	ACO	·	·	□	□	•	□	·	(40x100)	·	✓	3
21	Moghaddam et al.	2005	SA	·	·	·	·	•	·	·	(20x30)	·	✓	2
22	Prabhaharan et al.	2005	ACO	·	·	□	□	•	□	·	(40x100)	·	✓	
23	Venkumar and Haq	2005	NN	·	·	□	□	·	•	·	(40x100)	·	✓	
24	Andrés and Lozano	2006	PSO	·	·	□	□	•	□	·	(10x10)	•	✗	
25	Car and Mikac	2006	GA	·	·	·	·	•	·	·	(15x15)	•	✓	3
26	Defersha and Chen	2006	Hy	·	·	·	•	•	·	•	(30x90)	•	✗	
27	Nsakanda et al.	2006	H	·	·	□	□	•	□	·	(10000x25)	•	✗	
28	Foulds et al.	2006	TS	·	·	□	□	•	□	·	(5x7)	•	✗	
29	Lei and Wu	2006	TS	·	·	□	□	•	□	·	(15x30)	·	✓	
30	Nsakanda et al.	2006	Hy	·	·	·	·	•	·	•	(15x150)	•	✓	3
31	Torkul et al.	2006	NN	·	·	□	□	·	•	·	(18x25)	·	✓	
32	Venkumar and Haq	2006	NN	·	·	□	□	·	•	·	(40x100)	·	✓	
33	Won and Currie	2006	Pm	·	·	·	•	·	·	·	(41x30)	•	✓	1
34	James et al.	2007	Hy	·	·	·	·	•	·	•	(40x100)	•	✓	7
35	Saidi-Mehrabad and Safaei	2007	NN	·	·	□	□	·	•	·	(9x10)	•	✗	
36	Tavakkoli-Moghaddam et al.	2007	Hy	·	·	□	□	•	•	•	(10x10)	•	✗	

Table 4. Continued

No.	Source		CF method	Categories							Performance			
	Author(s)	Year		1	2	3	4	5	6	7	A	B	C	D
37	Wu et al.	2007	GA	·	·	□	□	•	□	·	NA	·	✗	
38	Boulif and Atif	2008	Hy	·	·	·	·	•	•	•	(20x24)	•	✓	2
39	Defersha and Chen	2008	GA	·	·	□	□	•	□	·	6x12)	•	✗	
40	Durán et al.	2008	SS	·	·	□	□	•	□	·	(12x12)	·	✗	
41	Kao and Li	2008	ACO	·	·	□	□	•	□	·	(50x150)	•	✓	
42	Megala and Rajendran	2008	ACO	·	·	□	□	•	□	·	(40x100)	·	✓	
43	Papaioannou and Wilson	2008	F	·	·	□	□	·	•	·	(9x9)	·	✗	
44	Safaei et al.	2008	F	·	·	□	□	·	•	·	(6x8)	•	✗	
45	Safaei et al.	2008	Hy	·	·	·	•	·	•	•	(6x8)	•	✓	1
46	Spiliopoulos and Sofi-anopoulou	2008	ACO	·	·	·	·	•	·	·	(37x53)	·	✓	
47	Tavakkoli-Moghaddam et al.	2008	SA	·	·	·	·	•	·	·	(17x30)	•	✗	
48	Wu et al.	2008	SA	·	·	·	·	•	·	·	(40x100)	•	✓	3
49	Yang and Yang	2008	NN	·	·	·	·	·	•	·	(46x105)	•	✗	
50	Ahi et al.	2009	Hy	·	•	·	·	·	•	•	(20x51)	•	✓	2
51	Bajestani et al.	2009	PSO	·	·	·	·	•	·	·	(9x10)	·	✓	
52	Bajestani et al.	2009	SS	·	·	·	·	•	·	·	(9x10)	•	✓	3
53	Mahdavi, et al.	2009	GA	·	·	·	·	•	·	·	(40x100)	•	✓	6
54	Oliveira et al.	2009	CA	·	•	·	·	·	·	·	(46x100)	•	✗	
55	Wu et al.	2009	Hy	·	·	·	·	•	·	•	(40x100)	•	✓	3
56	Wu et al.	2009	Hy	·	·	·	·	•	·	•	(50x120)	•	✓	8
57	Deljoo et al.	2010	GA	·	·	·	·	•	·	·	(10x10)	•	✗	
58	Li et al.	2010	ACO	·	·	·	·	•	·	·	(50x150)	•	✓	11
59	Naadimuthu et al.	2010	Hy	·	·	·	·	·	•	•	(6x5)	•	✗	
60	Neto and Filho	2010	Hy	·	·	·	·	•	·	•	(13x13)	·	✗	
61	Noktehdan et al.	2010	Hy	·	·	·	·	•	·	•	(37x53)	•	✓	2
62	Pailla et al.	2010	GA	·	·	·	·	•	·	·	(40x100)	•	✓	8
63	Wu et al.	2010	WFA	·	·	·	·	•	·	·	(50x100)	·	✓	2

Notes:
(I) Source list of contributors
(II) Acronym of CF approach
ACO = ant colony optimization CA = cluster analysis F = fuzzy logic GA = genetic algorithm H = classical heuristic.
Hy = hybrid NN = neural network Pm = p-median PSO = particle swarm optimisation SA = simulated annealing.
SS = search scatter TS = tabu search WFA = water flow-like algorithm.
(III) Major categories of production-oriented cell formation methods
1 = Descriptive methods 2 = Cluster analysis 3 = Graph partitioning 4 = Mathematical programming.
5 = Heuristic methods 6 = Artificial intelligence 7 = Hybrid methods.
(IV) Performance and computational results of the CF methods used
A = max size of data set used (machines x parts) B = provides performance measure for max size of data set used.
C = comparison to other existing methods ('✓' = yes and '✗' = no) D = number of other existing methods used for comparison NA = not available.

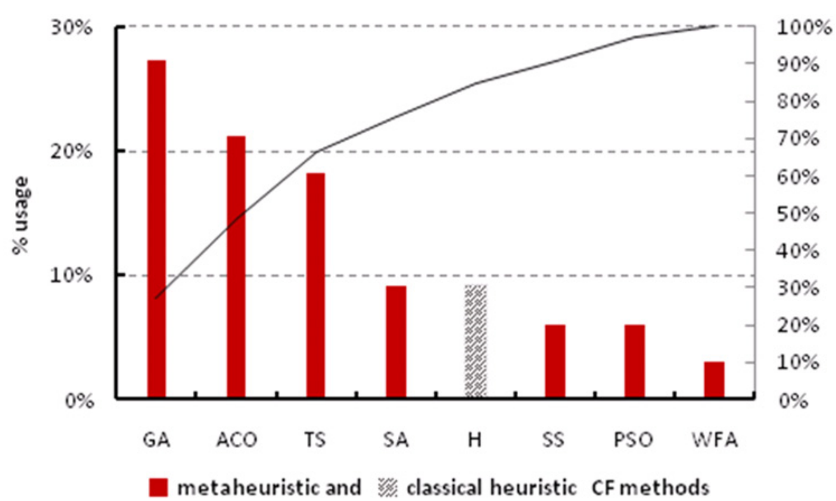
Figure 5. Frequency of individual categories for production-oriented CF methods



that CM research partially applies generic methods or techniques of POM.

From another point of view it is viable to identify causal relationships between concepts and tools in the two areas. This relationship can be seen, for example, between the JIT concept in the 1980s and an expansion of the heuristic methods in the 1980s. The extensive acceptance of the just-in-time (JIT) philosophy by various industries supported the development of cell manufacturing in firms as it is an important element in the successful implementation of just in time. With the rapidly increasing demand for such solutions, the need for more effective cell formation methods was naturally growing. This demand is likely to have accelerated the development of cell formation methods based on heuristics and mathematical programming.

Figure 6. The percentage usage of meta-heuristic and heuristic classical methods for CFP



Moreover, the surveys of POM and CM show some topical development directions, such as the importance of operation strategy in POM and the dominance of meta-heuristic techniques in cell formation problems.

REFERENCES

Ahi, A., Aryanezhad, M. B., Ashtiani, B., & Makui, A. (2009). A novel approach to determine cell formation, intracellular machine layout and cell layout in the CMS problem based on TOPSIS method. *Computers & Operations Research*, *36*(5), 1478–1496. doi:10.1016/j.cor.2008.02.012

Albadawi, Z., Bashir, H. A., & Chen, M. (2005). A mathematical approach for the formation of manufacturing cells. *Computers & Industrial Engineering*, *48*(1), 3–21. doi:10.1016/j.cie.2004.06.008

Anderson, D., Sweeney, D., & Williams, T. (2002). *An introduction to management science: Quantitative approaches to decision making* (10th ed.). Cincinnati, OH: South-Western Publishing Company.

Baber, Z. (1996). *The science of empire*. Albany, NY: State University of New York Press.

Bajestani, M. A., Rabbani, M., Rahimi-Vahed, A. R., & Khoshkhou, G. B. (2009). A multi-objective scatter search for a dynamic cell formation problem. *Computers & Operations Research*, *36*(3), 777–794. doi:10.1016/j.cor.2007.10.026

Bicheno, J., & Elliott, B. B. (1997). *Operations management: An active learning approach*. Oxford, UK: Blackwell Publishing.

Boctor, F. F. (1991). A linear formulation of the machine-part cell formation problem. *International Journal of Production Research*, *29*(2), 343–356. doi:10.1080/00207549108930075

Boulif, M., & Atif, K. (2008). A new fuzzy genetic algorithm for the dynamic bi-objective cell formation problem considering passive and active strategies. *International Journal of Approximate Reasoning*, *47*(2), 141–165. doi:10.1016/j.ijar.2007.03.003

Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: Catching the wave. *Harvard Business Review*, *73*(1), 43–53.

Bowman, E. H., & Feter, R. B. (1957). *Analysis for production management*. Homewood, IL: R. D. Irwin, Inc.

Bueno, E., Morcillo, P., & Rodriguez Anton, J. M. (1997). Management of technology: Proposal for a diagnosis model. *The Journal of High Technology Management Research*, *8*, 63–87. doi:10.1016/S1047-8310(97)90014-6

Buffa, E. S. (1961). *Modern production management*. New York, NY: Wiley.

Burbidge, J. L. (1971). Production flow analysis. *Production Engineering*, *50*, 139–152. doi:10.1049/tpe.1971.0022

Burbidge, J. L. (1977). A manual method of production flow analysis. *Production Engineering*, *56*(1), 34–38. doi:10.1049/tpe.1977.0129

Cao, D., & Chen, M. (2004). Using penalty function and Tabu search to solve cell formation problems with fixed cell cost. *Computers & Operations Research*, *31*(1), 21–37. doi:10.1016/S0305-0548(02)00144-2

Car, Z., & Mikac, T. (2006). Evolutionary approach for solving cell-formation problem in cell manufacturing. *Advanced Engineering Informatics*, *20*(3), 227–232. doi:10.1016/j.aei.2006.01.005

Caux, C., Bruniaux, R., & Pierreval, H. (2000). Cell formation with alternative process plans and machine capacity constraints: A new combined approach. *International Journal of Production Economics*, *64*(1-3), 279–28. doi:10.1016/S0925-5273(99)00065-1

Chan, H. M., & Milner, D. A. (1982). Direct clustering algorithm for group formation in cellular manufacturing. *Journal of Manufacturing Systems*, *1*, 65–74. doi:10.1016/S0278-6125(82)80068-X

Chandrasekharan, M. P., & Rajagopalan, R. (1986). MODROC: An extension of rank order clustering for group technology. *International Journal of Production Research*, *24*, 1221–1233. doi:10.1080/00207548608919798

Chandrasekharan, M. P., & Rajagopalan, R. (1987). ZODIAC: An algorithm for concurrent formation of part-families and machine-ceils. *International Journal of Production Research*, *25*(6), 835–850. doi:10.1080/00207548708919880

Chase, R. B., & Aquilano, N. J. (1989). *Production and operation management: A life cycle approach* (5th ed.). Homewood, IL: Richard D. Irwin, Inc.

Chase, R. B., Aquilano, N. J., & Jacobs, F. R. (2004). *Operations management for competitive advantage* (10th ed.). Boston, MA: McGraw-Hill.

Christensen, C. M. (2002). *The innovator's dilemma: The revolutionary book that will change the way you do business*. New York, NY: Harper Business Essential.

Corbett, C. J., & Van Wassehove, L. N. (1993). The natural drift: What happened to operations research? *Operations Research*, *41*(4), 625–640. doi:10.1287/opre.41.4.625

Deljoo, V., Al-e-Hashem, S. M. J. M., Deljoo, F., & Aryanezhad, M. B. (2010). Using genetic algorithm to solve dynamic cell formation problem. *Applied Mathematical Modelling*, *34*(4), 1078–1092. doi:10.1016/j.apm.2009.07.019

Drejer, A. (2002). Towards a model for contingency of management of technology. *Technovation*, *22*(6), 363–370. doi:10.1016/S0166-4972(01)00029-3

Fuller, J. A., & Martinec, C. L. (2005). Operations research and operations management: From selective optimization to system optimization. *Journal of Business & Economics Research*, *3*(7), 11–15.

Gonçalves, J. F., & Resende, M. G. C. (2004). An evolutionary algorithm for manufacturing cell formation. *Computers & Industrial Engineering*, *47*(2-3), 247–273. doi:10.1016/j.cie.2004.07.003

Guerrero, F., Lozano, S., Smith, K. A., Canca, D., & Kwok, T. (2002). Manufacturing cell formation using a new self-organizing neural network. *Computers & Industrial Engineering*, *42*(2-4), 377–382. doi:10.1016/S0360-8352(02)00039-6

Hambrick, D. C., & Frederickson, J. W. (2001). Are you sure you have a strategy? *The Academy of Management Executive*, *15*(4). doi:10.5465/AME.2001.5897655

Irani, S. A., Subramanian, S., & Allam, Y. S. (1999). Introduction to cellular manufacturing systems. In Irani, S. A. (Ed.), *Handbook of cellular manufacturing systems*. USA: John Wiley & Sons. doi:10.1002/9780470172476.ch

James, T. L., Brown, E. C., & Keeling, K. B. (2007). A hybrid grouping genetic algorithm for the cell formation problem. *Computers & Operations Research*, *34*(7), 2059–2079. doi:10.1016/j.cor.2005.08.010

Jones, G. (2005). *Multinationals and global capitalism: From the nineteenth to the twenty-first century*. New York, NY: Oxford University Press.

Kaparthi, S., & Suresh, N. C. (1992). Machine-component cell formation in group technology: A neural network approach. *International Journal of Production Research*, *30*, 1353–1367. doi:10.1080/00207549208942961

Kazmi, A. (2008). *Strategic management and business policy, Tata*. New Delhi, India: McGraw Hill Publishing Company Ltd.

Kiggundu, M. N., & Ji, S. (2008). Global growth companies in emerging economies: New champions, new challenges. *Journal of Business and Behavioral Sciences*, *19*(1), 70–90.

King, J. R. (1980). Machine-component grouping in production flow analysis: An approach using rank order clustering algorithm. *International Journal of Production Research*, *18*, 213–232. doi:10.1080/00207548008919662

Kusiak, A. (1987). The generalized group technology concept. *International Journal of Production Research*, *25*, 561–569. doi:10.1080/00207548708919861

Landes, D. (1998). *The wealth and poverty of nations: Why some are so rich and others so poor*. New York, NY: W.W. Norton.

Li, X., Baki, M. F., & Aneja, Y. P. (2010). An ant colony optimization metaheuristic for machine–part cell formation problems. *Computers & Operations Research*, *37*(12), 2071–2081. doi:10.1016/j.cor.2010.02.007

Luss, H., & Rosenwein, M. B. (1997). Operations research applications: Opportunities and accomplishments. *European Journal of Operational Research*, *97*(2), 220–244. doi:10.1016/S0377-2217(96)00194-4

Mahdavi, I., Paydar, M. M., Solimanpur, M., & Heidarzade, A. (2009). Genetic algorithm approach for solving a cell formation problem in cellular manufacturing. *Expert Systems with Applications*, *36*(3), 6598–6604. doi:10.1016/j.eswa.2008.07.054

Mansouri, S. A., Husseini, S. M., & Newman, S. T. (2000). A review of the modern approaches to multi-criteria cell design. *International Journal of Production Research*, *38*, 1201–1218. doi:10.1080/002075400189095

McAuley, J. (1972). Machine grouping for efficient production. *Production Engineering*, *51*(2), 53–57. doi:10.1049/tpe.1972.0006

McCormick, W. T., Scweitzer, P. J., & White, T. W. (1972). Problem decomposition and data reorganization by a cluster technique. *Operations Research*, *20*, 993–1009. doi:10.1287/opre.20.5.993

Mosier, C. T. (1989). An experiment investigating the application of clustering procedures and similarity coefficients to the GT machine cell formation problem. *International Journal of Production Research*, *27*(10), 1811–1835. doi:10.1080/00207548908942656

Naadimuthu, G., Gultom, P., & Lee, E. S. (2010). Fuzzy clustering in cell formation with multiple attributes. *Computers & Mathematics with Applications (Oxford, England)*, *59*(9), 3137–3147. doi:10.1016/j.camwa.2010.02.038

Neto, A. R. P., & Filho, E. V. G. (2010). A simulation-based evolutionary multiobjective approach to manufacturing cell formation. *Computers & Industrial Engineering*, *59*(1), 64–74. doi:10.1016/j.cie.2010.02.017

Noktehdan, A., Karimi, B., & Kashan, A. H. (2010). A differential evolution algorithm for the manufacturing cell formation problem using group based operators. *Expert Systems with Applications: An International Journal*, *37*(7), 4822–4829. doi:10.1016/j.eswa.2009.12.033

Nsakanda, A. L., Diaby, M., & Price, W. L. (2006). Hybrid genetic approach for solving large-scale capacitated cell formation problems with multiple routings. *European Journal of Operational Research*, *171*(3), 1051–1070. doi:10.1016/j.ejor.2005.01.017

Offodile, O. F., Mehrez, A., & Grznar, J. (1994). Cellular manufacturing: A taxonomic review framework. *Journal of Manufacturing Systems*, *13*(3), 196–220. doi:10.1016/0278-6125(94)90005-1

Oliveira, S., Ribeiro, J. F. F., & Seok, S. C. (2009). A spectral clustering algorithm for manufacturing cell formation. *Computers & Industrial Engineering*, *57*(3), 1008–1014. doi:10.1016/j.cie.2009.04.008

Onwubolu, G. C., & Mutingi, M. (2001). A genetic algorithm approach to cellular manufacturing systems. *Computers & Industrial Engineering*, *39*(1-2), 125–144. doi:10.1016/S0360-8352(00)00074-7

Pailla, A., Trindade, A. R., Parada, V., & Ochi, L. S. (2010). A numerical comparison between simulated annealing and evolutionary approaches to the cell formation problem. *Expert Systems with Applications*, *37*(7), 5476–5483. doi:10.1016/j.eswa.2010.02.064

Rajagopalan, R., & Batra, J. (1975). Design of cellular production systems - A graph theoretic approach. *International Journal of Production Research*, *13*, 56–68. doi:10.1080/00207547508943029

Safaei, N., Saidi-Mehrabad, M., Tavakkoli-Moghaddam, R., & Sassani, F. (2008). A fuzzy programming approach for a cell formation problem with dynamic and uncertain conditions. *Fuzzy Sets and Systems*, *159*(2), 215–236. doi:10.1016/j.fss.2007.06.014

Seifoddini, H., & Wolfe, P. M. (1986). Application of the similarity coefficient method in group technology. *IIE Transactions*, *18*(3), 271–277. doi:10.1080/07408178608974704

Selim, H. M., Askin, R. G., & Vakharia, A. J. (1998). Cell formation in group technology: Review, evaluation and direction for future research. *Computers & Industrial Engineering*, *34*, 2–30. doi:10.1016/S0360-8352(97)00147-2

Slack, N., & Lewis, M. (2002). *Operations strategy*. London, UK: Financial-Times Prentice Hall.

Soleymanpour, M., Vrat, P., & Shankar, R. (2002). A transiently chaotic neural network approach to the design of cellular manufacturing. *International Journal of Production Research*, *40*(10), 2225–2244. doi:10.1080/00207540210122284

Steers, R. M., & Nardon, L. (2006). *Managing in the global economy*. Armonk, NY: M. E. Sharpe, Inc.

Swink, M., & Way, M. H. (1995). Manufacturing strategy: Propositions, current research, renewed directions. *International Journal of Operations & Production Management*, *15*(7), 4–26. doi:10.1108/01443579510090381

Tavakkoli-Moghaddam, R., Aryanezhad, M. B., Safaei, N., Vasei, M., & Azaron, A. (2007). A new approach for the cellular manufacturing problem in fuzzy dynamic conditions by a genetic algorithm. *Journal of Intelligent and Fuzzy Systems*, *8*(4), 363–376.

Tavakkoli-Moghaddam, R., Safaei, N., & Sassani, F. (2008). A new solution for a dynamic cell formation problem with alternative routing and machine costs using simulated annealing. *The Journal of the Operational Research Society*, *23*, 916–924.

Taylor, F. W. (1911). *The principles of scientific management*. New York, NY: Harper & Brothers.

Thomas, A., & Grabot, B. (2006). *Key technologies and strategies for creating sustainable manufacturing organizations.* IPROMS 2006 - Virtual International Conference on Intelligent Production Machines and Systems, Manufacturing Engineering Centre (MEC) Cardiff University Queen's Buildings.

Turner, W. C., Mize, J. H., Case, K. E., & Nazemetz, J. W. (1993). *Introduction to industrial and systems engineering*. Englewood Cliffs, NJ: Prentice Hall.

Wemmerlöv, U., & Hyer, N. L. (1986). Procedures for partfamily/machine group identification problem in cellular manufacturing. *Journal of Operations Management*, *6*, 125–145. doi:10.1016/0272-6963(86)90021-5

Won, Y., & Currie, K. R. (2006). An effective p-median model considering production factors in machine cell/part family formation. *Journal of Manufacturing Systems*, *25*(1), 58–64. doi:10.1016/S0278-6125(06)80033-6

Won, Y., & Lee, K. C. (2004). Modified p-median approach for efficient GT cell formation. *Computers & Industrial Engineering*, *46*(3), 495–510. doi:10.1016/j.cie.2004.01.010

Wu, T., Chang, C., & Chung, S. (2008). A simulated annealing algorithm for manufacturing cell formation problems. *Expert Systems with Applications: An International Journal*, *34*(3), 1609–1617. doi:10.1016/j.eswa.2007.01.012

Wu, T. H., Chang, C. C., & Yeh, J. Y. (2009). A hybrid heuristic algorithm adopting both Boltzmann function and mutation operator for manufacturing cell formation problems. *International Journal of Production Economics*, *120*(2), 669–688. doi:10.1016/j.ijpe.2009.04.015

Wu, T. H., Chung, S. H., & Chang, C. C. (2009). Hybrid simulated annealing algorithm with mutation operator to the cell formation problem with alternative process routings. *Expert Systems with Applications*, *36*(2), 3652–3661. doi:10.1016/j.eswa.2008.02.060

Wu, T. H., Chung, S. H., & Chang, C. C. (2010). A water flow-like algorithm for manufacturing cell formation problems. *European Journal of Operational Research*, *205*(2), 346–360. doi:10.1016/j.ejor.2010.01.020

Wu, T.-H., Low, C., & Wu, W.-T. (2004). A tabu search approach to the cell formation problem. *International Journal of Advanced Manufacturing Technology*, *23*, 916–924. doi:10.1007/s00170-003-1766-z

Wu, X., Chao-Hsien, C., Wang, Y., & Yan, W. (2007). A genetic algorithm for cellular manufacturing design and layout. *European Journal of Operational Research*, *181*, 156–167. doi:10.1016/j.ejor.2006.05.035

Yang, M. S., & Yang, J. H. (2008). Machine-part cell formation in group technology using a modified ART1 method. *European Journal of Operational Research*, *188*(1), 140–152. doi:10.1016/j.ejor.2007.03.047

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 1-20, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 74
Fashion Supply Chain Management through Cost and Time Minimization from a Network Perspective

Anna Nagurney
University of Massachusetts Amherst, USA

Min Yu
University of Massachusetts Amherst, USA

ABSTRACT

In this chapter, we consider fashion supply chain management through cost and time minimization, from a network perspective, and in the case of multiple fashion products. We develop a multicriteria decision-making optimization model subject to multimarket demand satisfaction, and provide its equivalent variational inequality formulation. The model allows for the determination of the optimal multiproduct fashion flows associated with the supply chain network activities, in the form of: manufacturing, storage, and distribution, and identifies the minimal total operational cost and total time consumption. The model allows the decision-maker to weigh the total time minimization objective of the supply chain network for the time-sensitive fashion products, as appropriate. Furthermore, we discuss potential applications to fashion supply chain management through a series of numerical examples.

INTRODUCTION

In recent decades, fashion retailers, such as Benetton, H&M, Topshop, and Zara have revolutionized the fashion industry by following what has become known as the "fast fashion" strategy, in which retailers respond to shifts in the market within just a few weeks, versus an industry average of six months (Sull and Turconi (2008)). Specifically, fast fashion is a concept developed in Europe to serve markets for teenage and young adult women who desire trendy, short-cycle, and relatively inexpensive clothing, and who are willing to buy from small retail shops and boutiques

DOI: 10.4018/978-1-4666-1945-6.ch074

(Doeringer and Crean (2006)). Fast fashion chains have grown quicker than the industry as a whole and have seized market share from traditional rivals (Sull and Turconi (2008)), since they aim to obtain fabrics, to manufacture samples, and to start shipping products with far shorter lead times than those of the traditional production calendar (Doeringer and Crean (2006)).

Nordas, Pinali, and Geloso Grosso (2006) further argued that time is a critical component in the case of labor-intensive products such as clothing as well as consumer electronics, both examples of classes of products that are increasingly time-sensitive. They presented two case studies of the textile and clothing sector in Bulgaria and the Dominican Republic, respectively, and noted that, despite higher production costs than in China, their closeness to major markets gave these two countries the advantage of a shorter lead time that allowed them to specialize in fast fashion products. Interestingly and importantly, the authors also identified that lengthy, time-consuming administrative procedures for exports and imports reduce the probability that firms will even enter export markets for time-sensitive products.

Clearly, superior time performance must be weighed against the associated costs. Indeed, as noted by So (2000), it can be costly to deliver superior time performance, since delivery time performance generally depends on the available capacity and on the operating efficiency of the system. It is increasingly evident that, in the case of time-sensitive products, with fashion being an example par excellence, an appropriate supply chain management framework for such products must capture both the operational (and other) cost dimension as well as the time dimension.

For example, in the literature, the total order cycle time, which refers to the time elapsed in between the receipt of customer order until the delivery of finished goods to the customer, is considered an important measure as well as a major source of competitive advantage (see Bower and Hout (1988) and Christopher (1992)), directly influencing the customer satisfaction level (cf. Gunasekaran, Patel, and Tirtiroglu (2001) and Towill (1997)). Moreover, according to the survey of Gunasekaran, Patel, and McGaughey (2004), performance metrics for time issues associated with planning, purchasing, manufacturing, and delivery are consistently rated as important factors in supply chain management.

Conventionally, there have been several methodological approaches utilized for time-dependent supply chain management, including multiperiod dynamic programming and queuing theory (see, e.g., Guide Jr., Muyldermans, and Van Wassenhove (2005), Lederer and Li (1997), Palaka, Erlebacher, and Kropp (1998), So and Song (1998), So (2000), Ray and Jewkes (2004), and Liu, Parlar, and Zhu (2007)). However, according to the review by Goetschalckx, Vidal, and Dogan (2002), the paper by Arntzen et al. (1995) is the only one that has captured the time issue in the modeling and design of a global logistics system, with the expression of time consumption explicitly in the objective function.

In particular, Arntzen et al. (1995) applied the Global Supply Chain Model (GSCM) to the Digital Equipment Corporation so as to evaluate global supply chain alternatives and to determine the worldwide manufacturing and distribution strategies. In their mixed-integer linear programming model to minimize the weighted combination of total cost and activity days, the authors adopted a weighted activity time to measure activity days throughout the supply chain, which is the sum of processing times for each individual segment multiplied by the number of units processed or shipped through the link. However, we believe that the authors oversimplified the weighted activity time in assuming that the unit processing activity days are fixed, regardless of the facility capacities and the product flows. Also, in some other mathematical models dealing with time-sensitive demand, the lead time is used as the only indicator to differentiate the demand groups (see Cheong, Bhatnagar, and Graves (2004)). We note that Ferdows, Lewis, and Machuca (2004) recognized the nonlinear relationship between capacity and time in the context of the fashion industry and fast response with a focus on Zara and, hence, an appropriate model for fashion supply chain management must be able to handle such nonlinearities.

In this paper, we utilize a network economics approach to develop a mathematical model for fashion supply chain management that allows a firm to determine its cost-minimizing and time-minimizing multiproduct flows, subject to demand satisfaction at the demand markets, with the inclusion of an appropriate weight associated with time minimization. Hence, we utilize a multicriteria decision-making perspective. In addition, we allow the cost on each network link, be it one corresponding to manufacturing (or procurement), to transportation/shipment, and/or to storage, or to any other type of product processing, which may also include administrative processing associated with importing/exporting, to be an increasing function of the flow in order to capture the aspect of capacity and, in effect, congestion, as would result in queuing phenomena. Hence, we take some ideas from the transportation and logistics literature (cf. Nagurney (1999) and the references therein). Similar assumptions we impose on the link time functions since, clearly, the time to process a volume of fashion product should be dependent on the flow. Given the realities of the fashion industry in the US (see, e.g., Sen (2008)), it is imperative to have a methodological framework that can provide decision-makers with both cost and time information associated with the complex network of fashion supply chain activities. As early as Fisher (1997) it has been recognized that different products may require distinct supply chains.

Multicriteria decision-making for supply chain management applications has been applied in both centralized and decentralized decision-making contexts and in the case of general, multitiered networks (see, e.g., Nagurney (2006) and Nagurney and Qiang (2009) and the references therein) with the most popular criteria utilized being cost, quality, and on-time delivery (Ho, Xu and Dey (2010)). Nagurney et al. (2005), in turn, developed a multitiered competitive supply chain network equilibrium model with supply side and demand side risk (see also Dong et al. (2005) and Nagurney and Matsypura (2005)). Nagurney and Woolley (2010) studied the decision-making problem associated with supply chain network integration, in the context of mergers and acquisitions, so as to minimize the cost and the emissions generated. Nagurney and Nagurney (2010) added environmental concerns into a supply chain network design model. In this paper, we capture the explicit time consumption associated with fashion supply chain activities, along with the associated costs, within a network framework. The model in this paper provides decision-makers with insights associated with trade-offs between the operational costs and the time involved in a multiproduct fashion supply chain subject to multimarket demand satisfaction.

This paper is organized as follows. In ”The Fashion Supply Chain Management Model,” we develop the fashion supply chain management model and reveal the generality of the associated network framework. We provide both the multicriteria decision-making optimization model as well as its equivalent variational inequality formulation. The latter is given, for the sake of generality, since it provides us with the foundation to also develop models for multiproduct competition in the fashion industry, with results on supply chain network design under oligopolistic competition and profit maximization obtained in Nagurney (2010). In addition, the variational inequality form allows for the efficient and effective computation of the multiproduct supply chain network flows. We also provide some qualitative properties.

In “Numerical Examples” we illustrate the model and its potential applications to fashion supply chain management through a series of numerical examples. In the concluding section, we summarize the results in this paper and provide suggestions for future research.

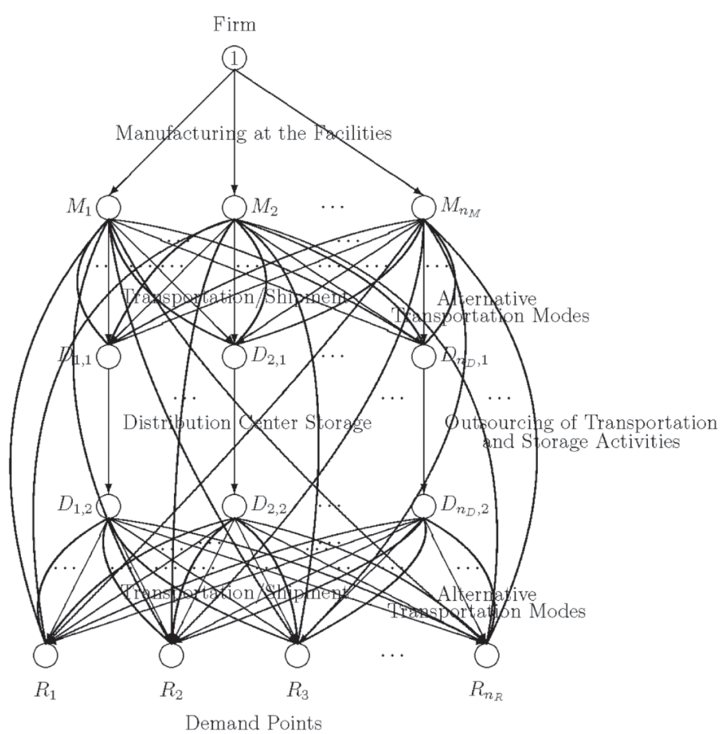
THE FASHION SUPPLY CHAIN MANAGEMENT MODEL

We assume that the fashion firm is involved in the production, storage, and distribution of multiple fashion products and is seeking to determine its optimal multiproduct flows to its demand points (markets) under total cost minimization and total time minimization, with the latter objective function weighted by the fashion firm.

We consider the fashion supply chain network topology depicted in Figure 1 but emphasize that the modeling framework developed here is not limited to such a network. This network is only representative, for definiteness. The origin node in the network in Figure 1 consists of node 1, which represents the beginning of the product processing, and the destination nodes, $R_1, \ldots, R_{n_R}$, are the demand points (markets) located at the bottom tier of the network. The paths joining the origin node to the destination nodes represent sequences of supply chain network activities corresponding to directed links that ensure that the fashion products are produced and, ultimately, delivered to the demand points. Hence, different supply chain network topologies to that depicted in Figure 1 correspond to distinct fashion supply chain network problems. For example, if the fashion product(s) can be delivered directly to the demand points from a manufacturing plant, then there would be, as depicted, links joining the corresponding nodes.

We assume that the fashion producing firm is involved in the production, storage, and transportation / distribution of *J* products, with a typical product denoted by *j*. In particular, as depicted in Figure 1, we assume that the firm has, at its dis-

Figure 1. The fashion supply chain network topology



posal, n_M manufacturing facilities/plants; n_D distribution centers, and must serve the n_R demand points. The links from the top-tiered node are connected to the manufacturing facility nodes of the firm, which are denoted, respectively, by: $M_1, \dots, M_{n_M}$. The links from the manufacturing facility nodes, in turn, are connected to the distribution/storage center nodes of the firm, which are denoted by $D_{1,1}, \dots, D_{n_D,1}$. Here we allow for the possibility of multiple links joining each such pair of nodes to reflect possible alternative modes of transportation/shipment between the manufacturing facilities and the distribution centers, an issue highly relevant to the fashion industry.

The links joining nodes $D_{1,1}, \dots, D_{n_D,1}$ with nodes $D_{1,2}, \dots, D_{n_D,2}$ correspond to the possible storage links for the products. Finally, there are multiple transportation/shipment links joining the nodes $D_{1,2}, \dots, D_{n_D,2}$ with the demand nodes: $R_1, \dots, R_{n_R}$. Distinct such links also correspond to different modes of transportation/shipment.

The outermost links in Figure 1 can also depict the option of possible outsourcing of the transportation and storage activities, with appropriate assigned costs and time values, as will be discussed below. Indeed, our supply chain network framework is sufficiently general and flexible to also capture alternatives (such as outsourcing of some of the supply chain network activities) that may be available to the fashion firm.

We assume that in the supply chain network topology there exists one path (or more) joining node 1 with each destination node. This assumption for the fashion supply chain network model guarantees that the demand at each demand point will be satisfied. We denote the supply chain network consisting of the graph $G=[N, L]$, where N denotes the set of nodes and L the set of directed links.

The demands for the fashion products are assumed as given and are associated with each product and each demand point. Let d_k^j denote the demand for the product j; $j=1,\dots,J$, at demand point R_k. A path consists of a sequence of links originating at the top node and denotes supply chain activities comprising manufacturing, storage, and transportation/shipment of the products to the demand nodes. Note that, if need be, one can also add other tiers of nodes and associated links to correspond to import/export administrative activities. Let x_p^j denote the nonnegative flow of product j on path p. Let P_k denote the set of all paths joining the origin node 1 with destination (demand) node R_k. The paths are assumed to be acyclic.

The following conservation of flow equations must hold for each product j and each demand point R_k:

$$\sum_{p \in P_k} x_p^j = d_k^j, \quad j = 1, \dots, J; \quad k = 1, \dots, n_R, \tag{1}$$

that is, the demand for each product must be satisfied at each demand point.

Links are denoted by a, b, etc. Let f_a^j denote the flow of product j on link a. We must have the following conservation of flow equations satisfied:

$$f_a^j = \sum_{p \in P} x_p^j \delta_{ap}, \quad j = 1, \dots, J; \quad \forall a \in L, \tag{2}$$

where $\delta_{ap}=1$ if link a is contained in path p and $\delta_{ap}=0$, otherwise. In other words, the flow of a product on a link is equal to the sum of flows of the product on paths that contain that link. Here P denotes the set of all the paths in Figure 1. The path flows must be nonnegative, that is,

$$x_p^j \geq 0, \quad j = 1, \dots, J; \quad \forall p \in P. \tag{3}$$

We group the path flows into the vector x and the link flows into the vector f, respectively.

Below we present the optimization problems in path flows and in link flows.

There is a unit operational cost associated with each product and each link (cf. Figure 1) of the network. We denote the unit cost on a link a associated with product j by c_a^j. . The unit cost of a link associated with each product, be it a manufacturing link, a transportation/shipment link, or a storage link, etc., is assumed, for the sake of generality, to be a function of the flow of all the products on the link. Hence, we have that

$$c_a^j = c_a^j(f_a^1,\ldots,f_a^J), \quad j=1,\ldots,J; \quad \forall a \in L. \tag{4}$$

Note that in the case of an outsourcing link for a fashion product the unit cost may be fixed, as per the negotiated contract.

Let C_p^j denote the unit operational cost associated with product j; j=1,…,J, on a path p, where

$$C_p^j = \sum_{a\in L} c_a^j \delta_{ap}, \quad j=1,\ldots,J; \quad \forall p \in P. \tag{5}$$

Then, the total operational cost for product j; j=1,…,J, on path p; p∈P, in view of (2), (4), and (5), can be expressed as:

$$\hat{C}_p^j(x) = C_p^j(x) \times x_p^j, \quad j=1,\ldots,J; \quad \forall p \in P. \tag{6}$$

The total cost minimization problem, hence, is formulated as:

$$\text{Minimize} \quad \sum_{j=1}^{J}\sum_{p\in P} \hat{C}_p^j(x), \tag{7}$$

subject to constraints (1) and (3).

In addition, the firm also seeks to minimize the time consumption associated with the demand satisfaction for each product at each demand point. Let t_a^j denote the average unit time consumption for product j; j=1,…,J, on link a, a∈L. We assume that

$$t_a^j = t_a^j(f_a^1,\ldots,f_a^J), \quad j=1,\ldots,J, \quad \forall a \in L, \tag{8}$$

that is, the link average unit time consumption is, also, for the sake of generality, a function of the flow of all the products on that link.

Therefore, the average unit time consumption for product j on path p is:

$$T_p^j = \sum_{a\in L} t_a^j \delta_{ap}, \quad j=1,\ldots,J, \quad \forall p \in P, \tag{9}$$

with the total time consumption for product j on path p, in view of (2), (8), and (9), given by:

$$\hat{T}_p^j(x) = T_p^j(x) \times x_p^j, \quad j=1,\ldots,J; \quad \forall p \in P. \tag{10}$$

The objective of time minimization problem is to minimize the total time associated with the supply chain network processing of all the products, which yields the following optimization problem:

$$\text{Minimize} \quad \sum_{j=1}^{J}\sum_{p\in P} \hat{T}_p^j(x), \tag{11}$$

subject to constraints (1) and (3).

The optimization problems (7) and (11) can be integrated into a single multicriteria objective function (cf. Dong et al. (2005)) using a weighting factor, ω, representing the preference of the decision-making authority. Please note that ω here can be interpreted as the monetary value of a unit of time. Consequently, the multicriteria decision–making problem, in path flows, can be expressed as:

$$\text{Minimize} \quad \sum_{j=1}^{J}\sum_{p\in P}\hat{C}_p^j(x)+\omega\sum_{j=1}^{J}\sum_{p\in P}\hat{T}_p^j(x), \tag{12}$$

subject to constraints (1) and (3).

The optimization problem (12) with the use of (2), (4), (5), (8), and (9), can be equivalently reformulated in link flows, rather than in path flows, as done above, as:

$$\text{Minimize} \quad \sum_{j=1}^{J}\sum_{a\in L}\hat{c}_a^j+\omega\sum_{j=1}^{J}\sum_{a\in L}\hat{t}_a^j, \tag{13}$$

subject to constraints (1)-(3) where $\hat{c}_a^j \equiv c_a^j(f_a^1,\ldots,f_a^J)\times f_a^j$ and the $\hat{t}_a^j \equiv t_a^j(f_a^1,\ldots,f_a^J)\times f_a^j$. We assume that the total link cost functions $\hat{c}_a^j$ and total time functions $\hat{t}_a^j$ are convex and continuously differentiable, for all products j and all links $a\in L$.

Let K denote the feasible set such that

$$K \equiv \{x \mid (1)\,\text{and}\,(3)\;\text{are satisfied}\}. \tag{14}$$

We now state the following result in which we derive the variational inequality formulations of the problem in both path flows and link flows, respectively. Having alternative formulations allows for the application of distinct algorithms (see, e.g., Nagurney (2006)).

THEOREM 1

A path flow vector $x^*\in K$ is an optimal solution to the optimization problem (12), subject to constraints (1) and (3), if and only if it is a solution to the variational inequality problem in path flows: determine the vector of optimal path flows, $x^*\in K$, such that:

$$\sum_{j=1}^{J}\sum_{p\in P}\left[\frac{\partial\hat{C}_p^j(x^*)}{\partial x_p^j}+w\,\frac{\partial\hat{T}_p^j(x^*)}{\partial x_p^j}\right]\times(x_p^j-x_p^{j^*})\geq 0,\quad \forall x\in K, \tag{15}$$

where

$$\frac{\partial\hat{C}_p^j(x)}{\partial x_p^j}\equiv\sum_{l=1}^{J}\sum_{a\in L}\frac{\partial\hat{c}_a^l(f_a^1,\ldots,f_a^J)}{\partial f_a^j}\delta_{ap},$$

and

$$\frac{\partial\hat{T}_p^j(x)}{\partial x_p^j}\equiv\sum_{l=1}^{J}\sum_{a\in L}\frac{\partial\hat{t}_a^l(f_a^1,\ldots,f_a^J)}{\partial f_a^j}\delta_{ap}.$$

A link flow vector $f^*\in K^1$ is an optimal solution to the optimization problem (13), subject to constraints (1) – (3), in turn, if and only if it is a solution to the variational inequality problem in link flows: determine the vector of optimal link flows, $f^*\in K^1$, such that:

$$\sum_{j=1}^{J}\sum_{l=1}^{J}\sum_{a\in L}\left[\frac{\partial\hat{c}_a^l(f_a^{1^*},\ldots,f_a^{J^*})}{\partial f_a^j}+\omega\,\frac{\partial\hat{t}_a^l(f_a^{1^*},\ldots,f_a^{J^*})}{\partial f_a^j}\right]\times(f_a^j-f_a^{j^*})\geq 0,\quad \forall f\in K^1, \tag{16}$$

where

$$K^1\equiv\{f \mid (1)-(3)\,\textit{are satisfied}\}.$$

Proof: The result follows from the standard theory of variational inequalities (see the book by Nagurney (1999) and the references therein) since the functions comprising the objective functions are convex and continuously differentiable under the imposed assumptions and the respective feasible sets consisting of the constraints are nonempty, closed, and convex. Q.E.D.

In addition, the following theoretical results in terms of the existence of solutions as well as the uniqueness of a link flow solution are immediate from the theory of variational inequalities. Indeed, the existence of solutions to (15) and (16) is guaranteed since the underlying feasible sets, K and K^1, are compact and the corresponding functions of marginal total costs and marginal total time are continuous, under the above assumptions. If the total link cost functions and the total time functions are strictly convex, then the solution to (16) is guaranteed to be unique.

It is worth noting that the above model contains, as a special case, the multiclass system-optimization transportation network model of Dafermos (1972) if we set ω=0. The fashion supply chain management network model developed here is novel since it captures both the reality of multiple products in this application domain as well as the significant relevant criteria of cost minimization as well as time minimization in the production and delivery of the fashion products to the demand markets.

Variational inequality (15) can be put into standard form (see Nagurney (1999)): determine $X^* \in$ K such that:

$$\langle F(X^*)^T, X - X^* \rangle \geq 0, \ \forall X \in \mathrm{K}, \tag{17}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in n-dimensional Euclidean space. Indeed, if we define the column vectors: X=x and

$$F(X) \equiv \begin{bmatrix} \dfrac{\partial \hat{C}_p^j(x)}{\partial x_p^j} + \omega \dfrac{\partial \hat{T}_p^j(x)}{\partial x_p^j}; \\ j = 1, \ldots, J; p \in P \end{bmatrix}, \tag{18}$$

and K=K then (15) can be re-expressed as (17).

Similarly, if we define the column vectors: X=f and

$$F(X) \equiv \begin{bmatrix} \dfrac{\partial \hat{c}_a^l(f_a^1, \ldots, f_a^J)}{\partial f_a^j} + \omega \dfrac{\partial \hat{t}_a^l(f_a^1, \ldots, f_a^J)}{\partial f_a^j}; \\ j = 1, \ldots, J; l = 1, \ldots, J; a \in L \end{bmatrix}, \tag{19}$$

and K=K^1 then (16) can be re-expressed as (17).

Note that the above model may be transformed into a single product network model by making as many copies of the network in Figure 1 as there are products and by constructing appropriate link total cost and time functions, which would be nonseparable, and by redefining the associated link flows, path flows, and demands accordingly. For details, see Nagurney and Qiang (2009) and the references therein.

NUMERICAL EXAMPLES

We now, for illustration purposes, present fashion supply chain numerical examples, both single product and multiproduct ones.

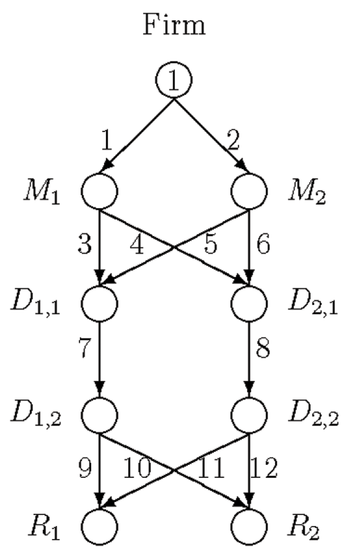
Single Product Fashion Supply Chain Examples

We assume that the fashion firm is involved in the production of a single fashion product and has, at its disposal, two manufacturing plants and two distribution centers. It must supply two different demand points. Hence, the topology is as depicted in Figure 2.

The manufacturing plant M_1 is located in the U.S., while the manufacturing plant M_2 is located off-shore and has lower operating cost. The average manufacturing time consumption of one unit of product is identical at these two plants, while the related costs vary mainly because of the different labor costs. The total cost functions and the total time functions for all the links are given in Table 1.

The demands for this fashion product at the demand points are:

Figure 2. The supply chain network topology for the numerical examples



d_1=100, d_2=200,

that is, the market at demand point R_1 is half that at demand market R_2.

We used the the general equilibration algorithm of Dafermos and Sparrow (1969) (see also, e.g., Nagurney (1999)) for the solution of the numerical examples.

We conducted sensitivity analysis by varying the value of time, ω, for ω=0,1,2,3,4,5. The computed optimal link flows are reported in Table 2.

We now display the optimal link flows as ω varies for the manufacturing links in Figure 3; for the first set of transportation links in Figure 4; for the set of storage links in Figure 5, and for the bottom tier of transportation links in Figure 6.

It is interesting to note from Figure 3 that, with the increase of the value of time, part of the fashion production is shifted from offshore manufacturing plant M_2 to onshore facility M_1, due to the onshore facility's advantage of shorter transportation time to distribution centers (or demand markets). Consequently, there is an increase in transportation flow from the onshore facility M_1 to the distribution centers, as depicted in Figure

Table 1. Total link operational cost and total time functions

Link *a*	$\hat{c}_a(f_a)$	$\hat{t}_a(f_a)$
1	$10f_1^2 + 10f_1$	$f_1^2 + 10f_1$
2	$f_2^2 + 5f_2$	$f_2^2 + 10f_2$
3	$f_3^2 + 3f_3$	$.5f_3^2 + 5f_3$
4	$f_4^2 + 4f_4$	$.5f_4^2 + 7f_4$
5	$2f_5^2 + 30f_5$	$.5f_5^2 + 25f_5$
6	$2f_6^2 + 20f_6$	$.5f_6^2 + 15f_6$
7	$.5f_7^2 + 3f_7$	$f_7^2 + 5f_7$
8	$f_8^2 + 3f_8$	$f_8^2 + 2f_8$
9	$f_9^2 + 2f_9$	$f_9^2 + 5f_9$
10	$2f_{10}^2 + f_{10}$	$f_{10}^2 + 3f_{10}$
11	$f_{11}^2 + 5f_{11}$	$f_{11}^2 + 2f_{11}$
12	$f_{12}^2 + 4f_{12}$	$f_{12}^2 + 4f_{12}$

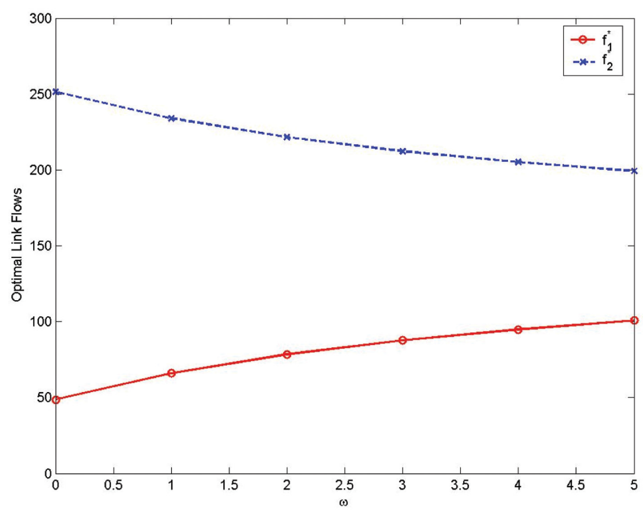
4. Figure 5, in turn, illustrates that distribution center D_2 is getting to be an appealing choice as the time performance concern increases, although the storage cost there is slightly higher than at D_1. Also, as the value of time increases, a volume of the fashion product flow switches from transportation link 9 (or link 12) to transportation link 11 (or link 10), to reduce the total time consumption of the distribution activities (as shown in Figure 6).

In Table 3, we provide the values of the total costs and the total time at the optimal solutions for the examples as ω increases.

Table 2. Computed optimal link flows f_a^* as ω increases

Link *a*	*ω*=0	*ω*=1	*ω*=2	*ω*=3	*ω*=4	*ω*=5
1	48.66	66.02	78.41	87.71	94.94	100.73
2	251.34	233.98	221.59	212.29	205.06	199.27
3	31.06	36.68	42.13	46.49	49.97	52.80
4	17.60	29.34	36.29	41.22	44.97	47.93
5	127.66	116.89	109.82	104.67	100.72	97.58
6	123.68	117.09	111.76	107.62	104.34	101.69
7	158.72	153.57	151.95	151.16	150.69	150.38
8	141.28	146.43	148.05	148.84	149.31	149.62
9	75.23	62.87	58.43	56.14	54.74	53.80
10	83.49	90.70	93.51	95.02	95.95	96.58
11	24.77	37.13	41.57	43.86	45.26	46.20
12	116.51	109.30	106.49	104.98	104.05	103.42

Figure 3. Optimal link flows on manufacturing links 1 and 2 as ω increases

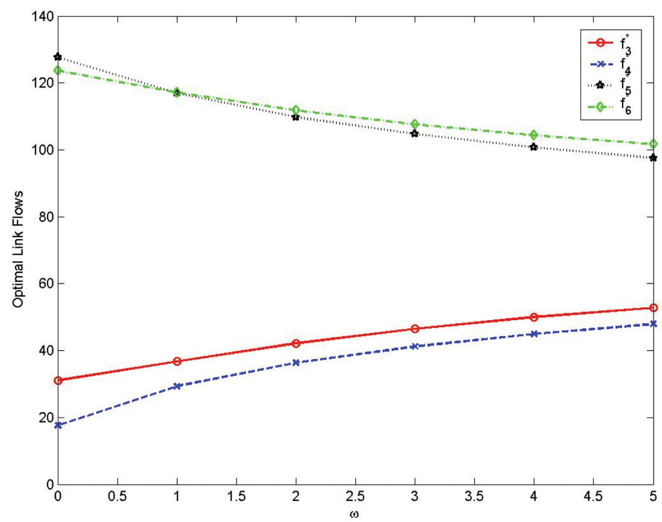


The values of the minimal total costs and the minimal total time for varying *ω* are displayed graphically in Figure 7. As can be seen from Figure 7, as the weight *ω* increases the minimal total time decreases, as expected, since a higher value of *ω* represents an increase in the decision-maker's valuation of time as a criterion.

Multiproduct Fashion Supply Chain Examples

We then considered multiproduct fashion supply chain problems. We assumed that the fashion firm provides two different fashion products with the same supply chain network topology as depicted in Figure 2. The total cost functions and the total time functions for all the links associated with

Figure 4. Optimal link flows on transportation links 3, 4, 5, and 6 as ω increases



product 1 and product 2 are given in Table 4 and 5, respectively.

The demands for the two fashion products at the demand points are:

$$d_1^1 = 100, \quad d_2^1 = 200,$$
$$d_1^2 = 300, \quad d_2^2 = 400.$$

To solves these problems, we used the modified projection method of Korpelevich (1977), embedded with the general equilibration algorithm of Dafermos and Sparrow (1969) (see also, e.g., Nagurney (1999)).

We also conducted sensitivity analysis, as in the section "Single Product Fashion Supply Chain Examples," by varying the value of time, ω, for

Figure 5. Optimal link flows on storage links 7 and 8 as ω increases

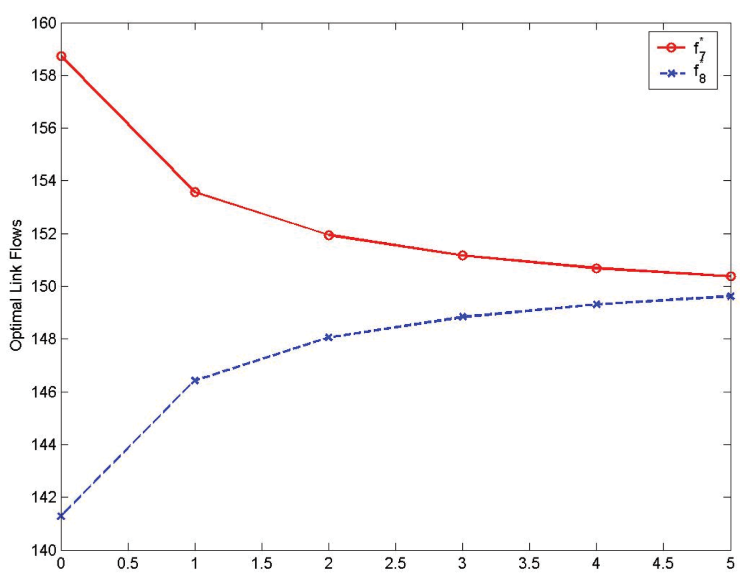


Figure 6. Optimal link flows on transportation links 9, 10, 11, and 12 as ω increases

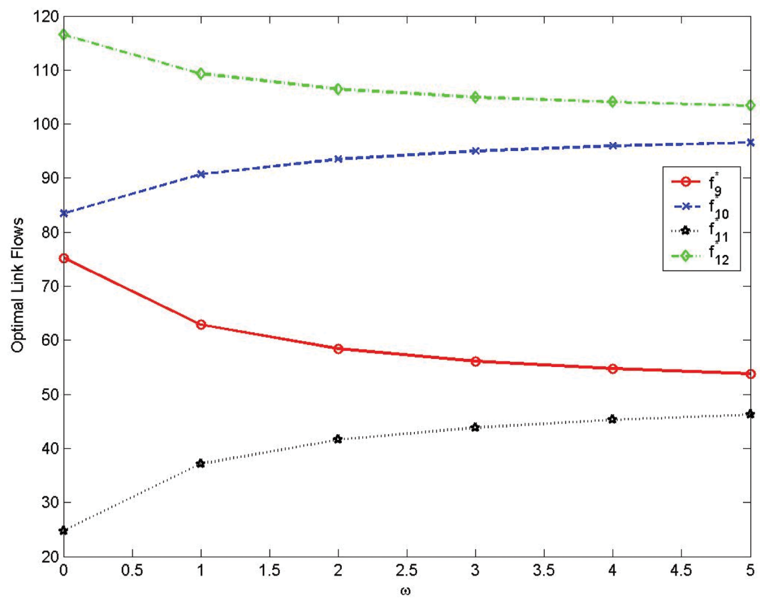


Table 3. Total costs and total times as ω increases

	ω=0	ω=1	ω=2	ω=3	ω=4	ω=5
Total cost	227,590.89	231,893.93	239,656.04	247,949.30	255,864.62	263,121.79
Total time	164,488.11	154,652.53	149,329.07	145,965.20	143,684.11	142,061.79

Figure 7. Minimal total costs and minimal total times as ω increases

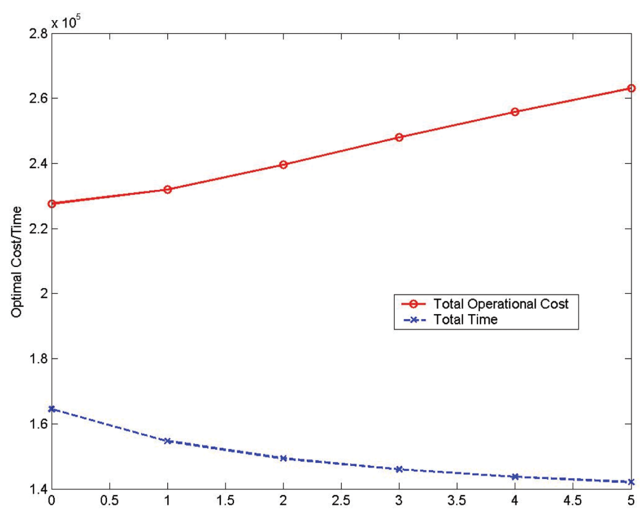


Table 4. Total link operational cost and total time functions for product 1

Link a	$\hat{c}_a^1(f_a^1, f_a^2)$	$\hat{t}_a^1(f_a^1, f_a^2)$
1	$10(f_1^1)^2 + 1f_1^1 f_1^2 + 10f_1^1$	$1(f_1^1)^2 + .3f_1^1 f_1^2 + 10f_1^1$
2	$1(f_2^1)^2 + .4f_2^1 f_2^2 + 5f_2^1$	$1(f_2^1)^2 + .3f_2^1 f_2^2 + 10f_2^1$
3	$1(f_3^1)^2 + .3f_3^1 f_3^2 + 3f_3^1$	$.5(f_3^1)^2 + .2f_3^1 f_3^2 + 5f_3^1$
4	$1(f_4^1)^2 + .2f_4^1 f_4^2 + 4f_4^1$	$.5(f_4^1)^2 + .2f_4^1 f_4^2 + 7f_4^1$
5	$2(f_5^1)^2 + .25f_5^1 f_5^2 + 30f_5^1$	$.5(f_5^1)^2 + .1f_5^1 f_5^2 + 25f_5^1$
6	$2(f_6^1)^2 + .3f_6^1 f_6^2 + 20f_6^1$	$.5(f_6^1)^2 + .1f_6^1 f_6^2 + 15f_6^1$
7	$.5(f_7^1)^2 + .1f_7^1 f_7^2 + 3f_7^1$	$1(f_7^1)^2 + .5f_7^1 f_7^2 + 5f_7^1$
8	$1(f_8^1)^2 + .1f_8^1 f_8^2 + 3f_8^1$	$1(f_8^1)^2 + .5f_8^1 f_8^2 + 2f_8^1$
9	$1(f_9^1)^2 + .5f_9^1 f_9^2 + 2f_9^1$	$1(f_9^1)^2 + .2f_9^1 f_9^2 + 5f_9^1$
10	$2(f_{10}^1)^2 + .3f_{10}^1 f_{10}^2 + 1f_{10}^1$	$1(f_{10}^1)^2 + .4f_{10}^1 f_{10}^2 + 3f_{10}^1$
11	$1(f_{11}^1)^2 + .6f_{11}^1 f_{11}^2 + 5f_{11}^1$	$1(f_{11}^1)^2 + .25f_{11}^1 f_{11}^2 + 2f_{11}^1$
12	$1(f_{12}^1)^2 + .7f_{12}^1 f_{12}^2 + 4f_{12}^1$	$1(f_{12}^1)^2 + .25f_{12}^1 f_{12}^2 + 4f_{12}^1$

ω=0,1,2,3,4,5. The computed optimal link flows associated with products 1 and 2 are, respectively, reported in Tables 6 and 7.

We display the optimal link flows of products 1 and 2 as ω varies for the manufacturing links in Figure 8; for the first set of transportation links in Figure 9; for the set of storage links in Figure 10, and for the bottom tier of transportation links in Figure 11.

With the increase of the value of time, parts of the production of fashion products 1 and 2 are shifted from offshore manufacturing plant M_2 to onshore facility M_1 (as depicted in Figure 8), resulting in an increase in transportation flow from M_1 to the distribution centers for both fashion products (as shown in Figure 9). However, Figure 10 illustrates that the distribution center D_2 is getting to be appealing for product 1 as the value of time increases, while the distribution center D_1 becomes attractive for product 2, since the distribution center D_1 is more time-efficient for product 2. In Figure 11, as the time performance concern increases, a volume of fashion product 1 switches from transportation link 9 to link 11; in contrast, the volume of flow of fashion product 2 on link 9 increases. Also, a volume of fashion product 2 switches from link 12 to link 10, while the flows of fashion product 1 on link 10 and 12 change slightly.

Table 5. Total link operational cost and total time functions for product 2

Link a	$\hat{c}_a^2(f_a^1, f_a^2)$	$\hat{t}_a^2(f_a^1, f_a^2)$
1	$8(f_1^2)^2 + 1f_1^1 f_1^2 + 10f_1^2$	$1(f_1^2)^2 + .5f_1^1 f_1^2 + 8f_1^2$
2	$1(f_2^2)^2 + .5f_2^1 f_2^2 + 4f_2^2$	$1(f_2^2)^2 + .5f_2^1 f_2^2 + 8f_2^2$
3	$1.5(f_3^2)^2 + .2f_3^1 f_3^2 + 3f_3^2$	$1(f_3^2)^2 + .1f_3^1 f_3^2 + 3f_3^2$
4	$1(f_4^2)^2 + .3f_4^1 f_4^2 + 4f_4^2$	$1(f_4^2)^2 + .2f_4^1 f_4^2 + 3f_4^2$
5	$2(f_5^2)^2 + .3f_5^1 f_5^2 + 25f_5^2$	$.8(f_5^2)^2 + .1f_5^1 f_5^2 + 20f_5^2$
6	$3(f_6^2)^2 + .4f_6^1 f_6^2 + 20f_6^2$	$.8(f_6^2)^2 + .2f_6^1 f_6^2 + 12f_6^2$
7	$1(f_7^2)^2 + .1f_7^1 f_7^2 + 3f_7^2$	$1(f_7^2)^2 + .4f_7^1 f_7^2 + 4f_7^2$
8	$.5(f_8^2)^2 + .2f_8^1 f_8^2 + 3f_8^2$	$1(f_8^2)^2 + .6f_8^1 f_8^2 + 4f_8^2$
9	$2(f_9^2)^2 + .3f_9^1 f_9^2 + 2f_9^2$	$1(f_9^2)^2 + .1f_9^1 f_9^2 + 7f_9^2$
10	$1(f_{10}^2)^2 + .5f_{10}^1 f_{10}^2 + 1f_{10}^2$	$1(f_{10}^2)^2 + .3f_{10}^1 f_{10}^2 + 6f_{10}^2$
11	$2(f_{11}^2)^2 + .5f_{11}^1 f_{11}^2 + 8f_{11}^2$	$1(f_{11}^2)^2 + .3f_{11}^1 f_{11}^2 + 3f_{11}^2$
12	$1(f_{12}^2)^2 + .4f_{12}^1 f_{12}^2 + 7f_{12}^2$	$1(f_{12}^2)^2 + .5f_{12}^1 f_{12}^2 + 4f_{12}^2$

The values of the total costs and the total time at the optimal solutions for the examples as ω increases are provided in Table 8, and displayed graphically in Figure 12. As expected, the minimal total time decreases as ω increases.

SUMMARY AND CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

In this paper, we developed a fashion supply chain management model, using a network economics perspective, that allows for multiple fashion products. The model consists of two objective functions: total cost minimization, associated with supply chain network activities, in the form of: manufacturing, storage, and distribution, and total time consumption minimization. A weighted objective function was then constructed with the weighting factor, representing the monetary value of a unit of time, decided by the firm.

We also provided the optimization model's equivalent variational inequality formulation, with nice features for computational purposes. The solution of the model yields the optimal multiproduct fashion flows of supply chain network activities, with the demands being satisfied at the minimal total cost and the minimal total time consumption. The model is illustrated with a spectrum of

Table 6. Computed optimal link flows f_a^{1*} as ω increases for product 1

Link *a*	ω=0	ω=1	ω=2	ω=3	ω=4	ω=5
1	53.81	71.31	83.38	92.24	99.02	104.39
2	246.19	228.69	216.62	207.76	200.98	195.61
3	41.58	43.44	47.03	50.23	52.90	55.10
4	12.23	27.87	36.35	42.01	46.13	49.29
5	125.73	114.20	106.97	101.86	98.02	95.01
6	120.45	114.49	109.64	105.90	102.96	100.60
7	167.32	157.64	154.00	152.09	150.91	150.11
8	132.68	142.36	146.00	147.91	149.09	149.89
9	73.27	63.60	60.07	58.25	57.14	56.39
10	94.05	94.04	93.93	93.84	93.78	93.73
11	26.73	36.40	39.93	41.75	42.86	43.61
12	105.95	105.96	106.07	106.16	106.22	106.27

Table 7. Computed optimal link flows f_a^{2*} as ω increases for product 2

Link *a*	ω=0	ω=1	ω=2	ω=3	ω=4	ω=5
1	150.04	192.14	219.13	237.85	251.58	262.07
2	549.96	507.86	480.87	462.15	448.42	437.93
3	17.38	63.50	87.28	102.35	112.87	120.65
4	132.66	128.64	131.85	135.50	138.71	141.42
5	309.57	278.09	259.11	246.45	237.40	230.61
6	240.39	229.77	221.75	215.70	211.02	207.32
7	326.95	341.59	346.39	348.80	350.27	351.26
8	373.05	358.41	353.61	351.20	349.73	348.74
9	141.99	145.71	147.19	148.00	148.52	148.88
10	184.96	195.88	199.20	200.80	201.75	202.38
11	158.01	154.29	152.81	152.00	151.48	151.12
12	215.04	204.12	200.80	199.20	198.25	197.62

numerical examples with potential application to fashion supply chain management.

The fashion supply chain network model allows the cognizant decision-maker to evaluate the effects of changes in the demand for its products on the total operations costs and time. It allows for the evaluation of changes in the cost functions and the time functions on total supply chain network costs and time. In addition, the flexibility of the network framework allows for the evaluation of the addition

Figure 8. Optimal link flows on manufacturing links 1 and 2 as ω increases

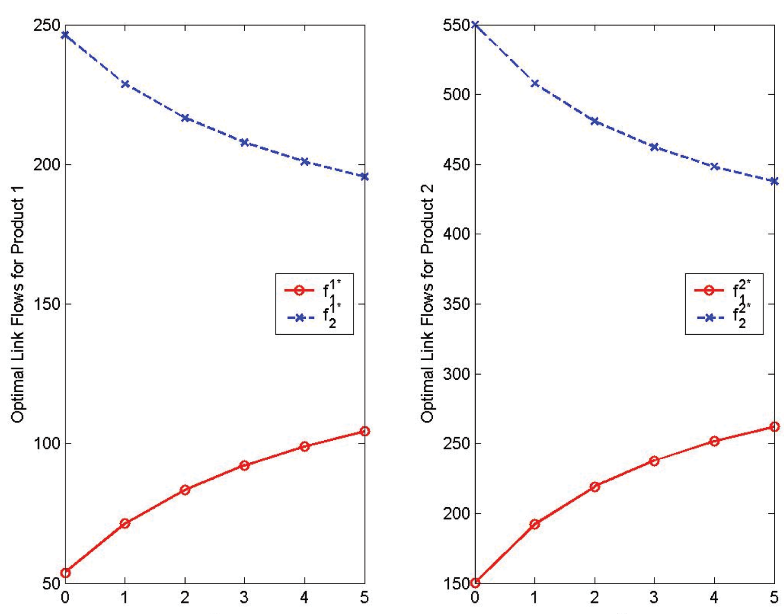
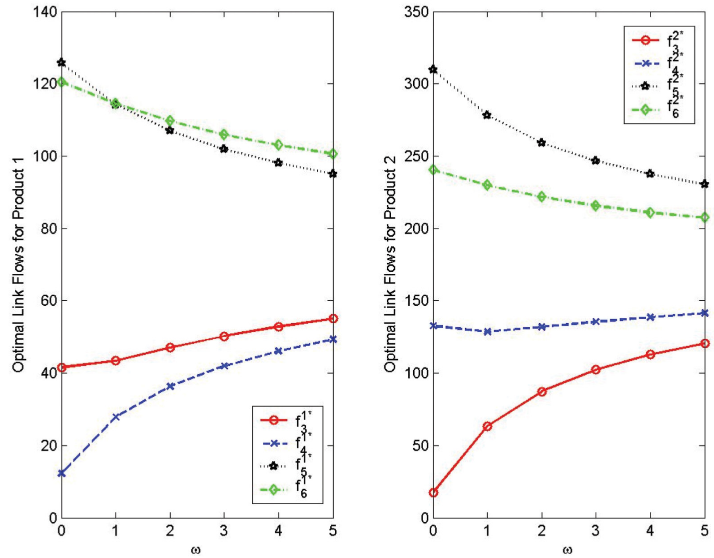


Figure 9. Optimal link flows on transportation links 3, 4, 5, and 6 as ω increases



of various links (or their removal) on the values of the objective function(s). Finally, the model, since it is network-based, is visually graphic.

The research in this paper can be extended in several directions. One can construct a fashion supply chain management model with price-sensitive and time-sensitive demands under oligopolistic competition. One can also incorporate environmental concerns and associated trade-offs. In addition, one can explore computationally as

Figure 10. Optimal link flows on storage links 7 and 8 as ω increases

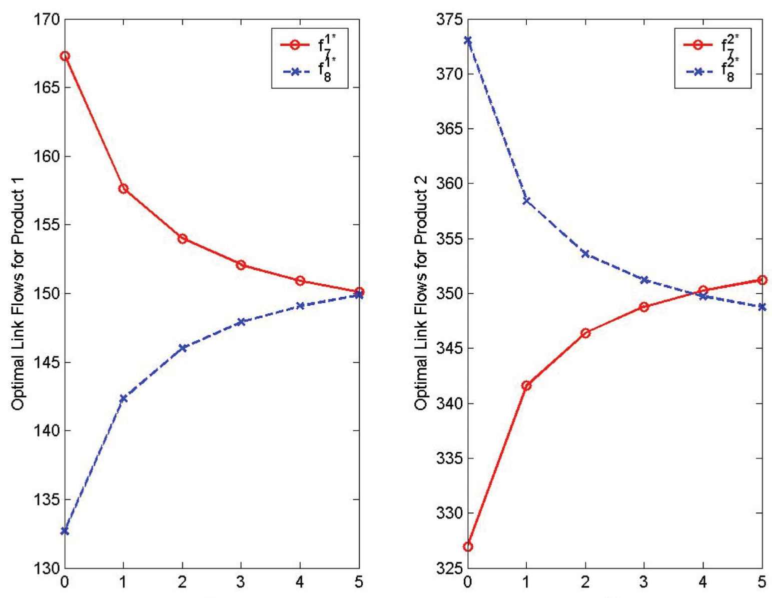


Figure 11. Optimal link flows on transportation links 9, 10, 11, and 12 as ω increases

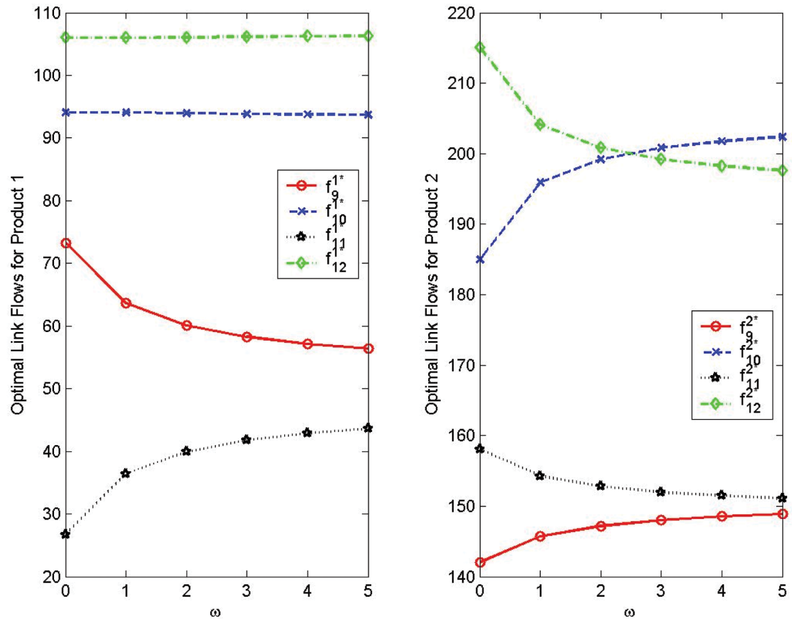


Figure 12. Minimal total costs and minimal total times as ω increases

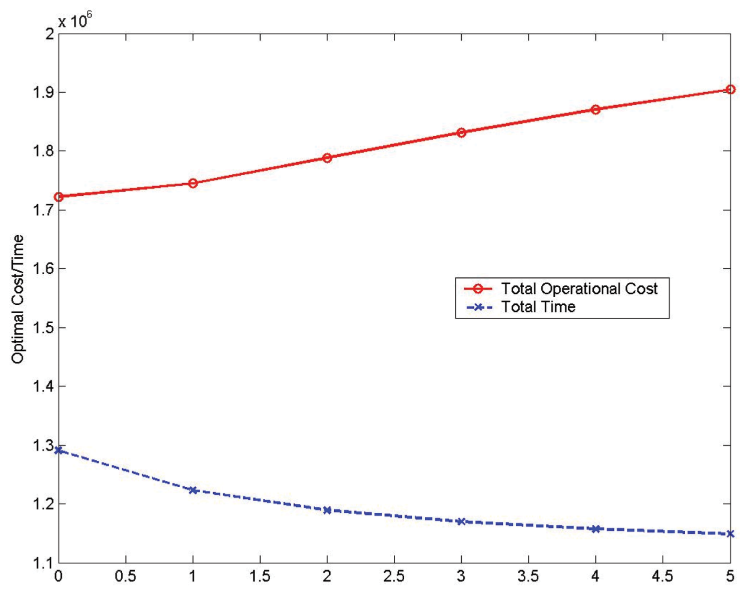


Table 8. Total costs and total times as ω increases

	ω=0	ω=1	ω=2	ω=3	ω=4	ω=5
Total cost	1,722,082.05	1,745,201.77	1,788,457.21	1,831,689.80	1,870,523.21	1,904,398.75
Total time	1,291,094.62	1,222,959.73	1,189,192.37	1,169,656.19	1,157,297.60	1,148,975.34

well as empirically large-scale fashion supply chain networks within our modeling framework. We leave such research for the future.

ACKNOWLEDGMENT

The authors acknowledge the helpful comments and suggestions of two anonymous reviewers. The authors also thank Professor Jason Choi for the opportunity to contribute to his edited volume. This research was supported, in part, by the John F. Smith Memorial Fund. This support is gratefully acknowledged.

REFERENCES

Arntzen, B. C., Brown, G. G., Harrison, T. P., & Trafton, L. L. (1995). Global supply chain management at Digital Equipment Corporation. *Interfaces*, *25*(1), 69–93. doi:10.1287/inte.25.1.69

Bower, J. L., & Hout, T. M. (1988). Fast cycle capability for competitive power. *Harvard Business Review*, *88*(6), 110–118.

Cheong, M. L. F., Bhatnagar, R., & Graves, S. C. (2004). *Logistics network deign with differentiated delivery lead-time: Benefits and insights*. Retrieved on June 8, 2010, from http://web.mit.edu/sgraves/www/papers/ Cheong1119.pdf

Christopher, M. (1992). *Logistics and supply chain management*. London, UK: Pitman Publishing.

Dafermos, S. C. (1972). The traffic assignment problem for multiclass-user transportation networks. *Transportation Science*, *6*(1), 73–87. doi:10.1287/trsc.6.1.73

Dafermos, S. C., & Sparrow, F. T. (1969). The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards*, *73B*, 91–118.

Doeringer, P., & Crean, S. (2006). Can fast fashion save the U.S. apparel industry? *Socio-economic Review*, *4*(3), 353–377. doi:10.1093/ser/mwl014

Dong, J., Zhang, D., Yan, H., & Nagurney, A. (2005). Multitiered supply chain networks: Multicriteria decision-making under uncertainty. *Annals of Operations Research*, *135*(1), 155–178. doi:10.1007/s10479-005-6239-3

Ferdows, K., Lewis, M. A., & Machuca, J. A. D. (2004). Rapid-fire fulfillment. *Harvard Business Review*, *82*(11), 104–110.

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review*, *75*(2), 105–116.

Goetschalckx, M., Vidal, C. J., & Dogan, K. (2002). Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. *European Journal of Operational Research*, *143*(1), 1–18. doi:10.1016/S0377-2217(02)00142-X

Guide, V. D. R. Jr, Muyldermans, L., & Van Wassenhove, L. N. (2005). Hewlett-Packard Company unlocks the value potential from time-sensitive returns. *Interfaces*, *35*(4), 281–293. doi:10.1287/inte.1050.0145

Gunasekaran, A., Patel, C., & McGaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, *87*(3), 333–347. doi:10.1016/j.ijpe.2003.08.003

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management*, *21*(1/2), 71–87. doi:10.1108/01443570110358468

Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, *202*(1), 16–24. doi:10.1016/j.ejor.2009.05.009

Korpelevich, G. M. (1977). The extragradient method for finding saddle points and other problems. *Matekon: Translations of Russian & East European Mathematical Economics*, *13*, 35–49.

Lederer, P. J., & Li, L. (1997). Pricing, production, scheduling, and delivery-time competition. *Operations Research*, *45*(3), 407–420. doi:10.1287/opre.45.3.407

Liu, L., Parlar, M., & Zhu, S. X. (2007). Pricing and lead time decisions in decentralized supply chains. *Management Science*, *53*(5), 713–725. doi:10.1287/mnsc.1060.0653

Nagurney, A. (1999). *Network economics: A variational inequality approach, second and* (revised edition). Boston, MA: Kluwer Academic Publishers.

Nagurney, A. (2006). *Supply chain network economics: Dynamics of prices, flows, and profits*. Cheltenham, England: Edward Elgar Publishing.

Nagurney, A. (2010). Supply chain network design under profit maximization and oligopolistic competition. *Transportation Research Part E, Logistics and Transportation Review*, *46*(3), 281–294. doi:10.1016/j.tre.2009.11.002

Nagurney, A., Cruz, J., Dong, J., & Zhang, D. (2005). Supply chain networks, electronic commerce, and supply side and demand side risk. *European Journal of Operational Research*, *164*(1), 120–142. doi:10.1016/j.ejor.2003.11.007

Nagurney, A., & Matsypura, D. (2005). Global supply chain network dynamics with multicriteria decision-making under risk and uncertainty. *Transportation Research Part E, Logistics and Transportation Review*, *41*(6), 585–612. doi:10.1016/j.tre.2005.07.002

Nagurney, A., & Nagurney, L. S. (2010). Sustainable supply chain network design: A multicriteria perspective. *International Journal of Sustainable Engineering*, *3*(3), 189–197. doi:10.1080/19397038.2010.491562

Nagurney, A., & Qiang, Q. (2009). *Fragile networks: Identifying vulnerabilities and synergies in an uncertain world*. Hoboken, NJ: John Wiley & Sons.

Nagurney, A., & Woolley, T. (2010). Environmental and cost synergy in supply chain network integration in mergers and acquisitions. In M. Ehrgott, B. Naujoks, T. Stewart, & J. Wallenius (Eds.), *Sustainable energy and transportation systems, Proceedings of the 19th International Conference on Multiple Criteria Decision Making, Lecture Notes in Economics and Mathematical Systems* (pp. 51-78). Berlin, Germany: Springer.

Nordas, H. K., Pinali, E., & Geloso Grosso, M. (2006). *Logistics and time as a trade barrier*. OECD Trade Policy Working Papers, No. 35. OECD Publishing. doi:10.1787/664220308873.

Palaka, K., Erlebacher, S., & Kropp, D. H. (1998). Lead-time setting, capacity utilization, and pricing decisions under lead-time dependent demand. *IIE Transactions*, *30*(2), 151–163. doi:10.1080/07408179808966447

Ray, S., & Jewkes, E. M. (2004). Customer lead time management when both demand and price are lead time sensitive. *European Journal of Operational Research*, *153*(3), 769–781. doi:10.1016/S0377-2217(02)00655-0

Sen, A. (2008). The US fashion industry: A supply chain review. *International Journal of Production Economics*, *114*(2), 571–593. doi:10.1016/j.ijpe.2007.05.022

So, K. C. (2000). Price and time competition for service delivery. *Manufacturing and Service Operations Management*, *2*(4), 392–409. doi:10.1287/msom.2.4.392.12336

So, K. C., & Song, J. S. (1998). Price, delivery time guarantees and capacity selection. *European Journal of Operational Research*, *111*(1), 28–49. doi:10.1016/S0377-2217(97)00314-7

Sull, D., & Turconi, S. (2008). Fast fashion lessons. *Business Strategy Review, Summer*, 5-11.

Towill, D. R. (1997). The seamless supply chain–The predator's strategic advantage. *International Journal of Technology Management*, *13*(1), 37–56. doi:10.1504/IJTM.1997.001649

This work was previously published in Fashion Supply Chain Management: Industry and Business Analysis, edited by Tsan-Ming Choi, pp. 1-20, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 75
An Exploratory Study on Product Lifecycle Management in the Fashion Chain:
Evidences from the Italian Leather Luxury Industry

Romeo Bandinelli
Università degli Studi di Firenze, Italy

Sergio Terzi
Università degli Studi di Bergamo, Italy

ABSTRACT

New Product Development (NPD) in most manufacturing sectors is stressed by an increasing global competition and pressure to improve product quality and innovation, reduce product cost and time-to-market (TTM), and rapidly respond to changing customer needs and shortened product lifecycles. These requirements are increasingly fulfilled by applying the PLM (Product Lifecycle Management) approach, a widely accepted concept that generally defines the adoption of a large number of ICT (Information and Communication Technology) solutions for managing product data along the product lifecycle. This contribution shows the results of research concerning the application of PLM within the luxury fashion supply chain, conducted in 2009 in Italy, with the analysis of 20 companies of the leather market. The research shows some of the differences that exist between the luxury industry and other more PLM-oriented sectors (e.g. automotive) in terms of adopted ICT tools, criticalities, problems, and benefits expected and realized.

DOI: 10.4018/978-1-4666-1945-6.ch075

INTRODUCTION

New Product Development (NPD) in most manufacturing sectors is stressed by an increasing global competition and pressure to improve product quality and innovation, reduce product cost and time-to-market (TTM), and rapidly respond to changing customer needs and shortened product lifecycles. These requirements are fulfilled by applying the PLM (Product Lifecycle Management) approach, a widely accepted concept that generally defines the adoption of a large number of ICT (Information and Communication Technology) solutions in order to support collaborative and coordinated environments, managing product data along the product lifecycle (from the cradle to the grave), and mainly focuses on the NPD process.

PLM has been extensively studied and analyzed in recent years, with different authors highlighting the holistic meaning of such acronym (e,g, Stark, Saaskvuoiri and Immonem, Grieves). For example, according to Stark (Stark, 2005), it brings together products, services, structures, activities, processes, people, skills, application systems, data, information, knowledge, techniques, practices and standards. Saaskvuori and Immonem (Saaskvuori and Immonem, 2005) define it as a comprehensive approach, entailing: (i) a strategic management perspective, wherein the product is the enterprise value creator; (ii) the application of a collaborative approach to better use the enterprise competences distributed amongst diverse business actors; and (iii) the adoption of plenty of ICT solutions in order to practically establish a coordinated, integrated and access-safe product information management environment in the extended context. Grieves (Grieves, 2006) identifies in PLM everything which deals with the management of the product data that are created, stored, and managed along the lifecycle of a product, from the design to end of life. Ma and Fuh (2008) state that PLM is a technical term to describe a comprehensive, systematic and scientific approach in managing enterprise performance based on a coherently and consistently integrated computer system that can effectively and efficiently fulfill the product and process information requirements within a dynamic, collaborative and networked environment.

PLM is one of the most recent evolutions of the enterprise ICT applications, characterized by an increasing integration and interoperability into and between tools and supported processes, both in design and management activities. In the last 30 years, CAx (Computer Aided, where x stands for Design, Styling, Manufacturing, Engineering, Process Planning, etc.) tools have successfully evolved from 2D to 3D modeling techniques, enabling a list of knowledge-based engineering approaches, generally addressed as Virtual Prototyping, Digital Mock-Up and Virtual Reality. Since the 1990's, collaboration in product design, development, and delivery has been enabled in integrated software platforms (alternatively addressed with new acronyms, like TDM Technical Data Management, EDM Engineering Data Management, PDM Product Data Management, PIM Product Information Management, etc.) where product data can be safely stored, retrieved, and reused. Since 2000, PLM has been used to define a comprehensive ICT market, nowadays composed of two main segments: (i) one constituted by a galaxy of design tools (i.e. CAx and engineering tools) and (ii) a second one composed by a plethora of Collaborative Product Definition/Development and Management platforms (CPDM). The first segment is dominated by ICT players historically dedicated to product design, like Autodesk, Dassault Systèmes, PTC, SIEMENS, while in the second these vendors compete with many other players coming from other backgrounds such as Oracle, SAP, etc.

PLM market is one of the most booming ICT markets irrespective of the global crisis, as many analysts (e.g. AMR Research, CIMdata, Daratech, Gartner) are continuously declaring, and PLM projects are considered as strategic investments, to be financed for surviving and preparing companies "for the next economy up-cycle" (Cimdata, 2008).

It has been widely applied in capital intensive industries such as aerospace and automotive. In such industries, the same PLM concept has been defined and it is still evolving, (Sapuan et al, 2006), (Tang and Qian, 2008), (Alemanni et al, 2008) becoming a comprehensive system supporting the information needs regarding products "from cradle to grave". This system involves many engineering stages such as industrial design, conceptual design, detailed design, process planning, manufacturing, assembly, sales, maintenance, and recycle or destroy (Ma and Fuh, 2008).

Today, aerospace, automotive, and similar mechanical-based sectors are considered traditional and conventional PLM markets in which PLM solutions (both CAx and CPDM) are well known and applied. In recent years, analysts and consulting companies have dedicated their attention to unconventional PLM sectors such as food, pharmaceutical, telecommunications, textiles, footwear, and apparel. Among these, the so-called "fashion luxury sector" has been addressed repeatedly as one of "the next" PLM sectors (Barret et al, 2007), (Sulesky et al, 2007), (Hojlo and Suleski, 2008).

Italy is commonly considered to be one of the most important "fashion and luxury" countries of the world (Brun et al, 2008), particularly with regard to the Italian leather industry. Within this context, the present contribution aims to investigate the declination of the PLM approach in the unconventional "fashion luxury sector", focusing on the application of PLM within the niche of the Italian leather sector, in order to understand the main commonalities and differences with more conventional PLM sectors.

Before presenting the results of the Italian-based research conducted by the authors in 2009, a brief description of this particular sector is reported. For supporting the reader, the Table 1 synthesizes the list of the acronyms adopted in the chapter.

BACKGROUND

The term "fashion" is used by practitioners and academics to generally refer to an industry which widely includes several sectors: textile, clothing, leather, knitwear, accessories, sunglasses, cosmetics, and jewelry. Each of these sectors show distinctive characteristics and may be further divided into different competitive segments characterized by a company's dimensions such as the served customers and markets, and applied technologies. Authors Saviolo and Testa (Saviolo and Testa, 2005) classify the fashion markets using the price and the number of sold units as variables that discriminate the type of the product, thus defining five main market segments: *couture*, *prêt-à-porter*, *diffusion*, *bridge*, and *mass*. From the last segment (*mass*) to the first one (*couture*) (i) the prize of the sold product increases, (ii) the number of sold units decreases, (iii) the perceived and the real quality increases, (iv) the product

Table 1. List of used acronyms

Definition	Acronym
Bill Of Material	BOM
Collaborative Product Definition / Development and Management	CPDM
Computer Aided Design	CAD
Computer Aided Manufacturing	CAM
Computer Aided x (where x stays for Technology, Design, Engineering, etc.)	CAx
Carry Over	CO
Engineering Data Management	EDM
Information and Communication Technology	ICT
New Product Development	NPD
Product Data Management	PDM
Product Information Management	PIM
Product Lifecycle Management	PLM
Small and Medium Enterprise	SME
Stock Keeping Units	SKU
Technical Data Management	TDM
Time To Market	TTM

differentiation and variety increases (from *mass* to *custom*) and (v) the demand predictability and stability generally decreases.

Each market segment has an impact in the structure of the fashion value chain in terms of TTM, lot dimensions, and number of sold units per product type. The value chain of the fashion industry has been studied by several researchers in the past years, and numerous publications can be found on such topic. Most of them are related to the fashion industry as general market (Brun et al, 2008) (Caniato et al, 2007), while some others are focused on textile and clothing (Burresi and Ravagni, 2006) (La Marca and Palamara, 2005) (Gereffi and Memedovic, 2003) (Richardson, 1996) (Bruce et al, 2004) (De Toni and Meneghetti, 2000). There are other papers as well focused on the description of specific value chain such as Zara (Ghemewat and Nueno, 2003) and (Vona, 2004), Burberry (Moore and Birtwistle, 2004) and Armani (Fernie et al, 1997). Saviolo and Testa (Saviolo and Testa, 2005) identify four types of fashion value chain according to the existing level of integration:

i. Integrated Griffe (e.g. Gucci, Prada, Ferragamo) companies which own and control nearly their entire value chains - from the design, to the production plants, until their branded shops.
ii. Industrial Integrated Groups (e.g. Tod's and Ermenegildo Zegna) which design, own their plants, and sometimes manage part of the distribution chain.
iii. Pure Brand Owners (e.g. Roberto Cavalli) which are generally small companies focused purely on design without involvement in the production, distribution, and selling processes.
iv. Industrial Players (e.g. Sergio Rossi) which provide their specific production and distribution competences and facilities (e.g. in the shoes making) to other brand owners.

Traditionally, a polarization can be observed in the fashion industry (Brun et al, 2008). Some companies choose to operate in the mass market, selling low-cost products that are available to a large number of consumers, whilst on the other hand other companies provide exclusive and expensive products to an elitist segment of consumers. The latter type composes the "luxury" segment. Companies which belong to both markets are not common. Like fashion, luxury itself is a transversal concept which dominates a plethora of manufacturing sectors including clothing, jewelry, furniture, wine making, automotive, and the yacht industries. Luxury is "a way of live" in which the customers search for elitist emotions. In the last decade, different authors (Fernie et al, 1997), (Silverstein et al, 2003) and (Sen, 2008) analyzed the intangible sense of elitism and desirability which pervade the luxury segment. For example, D'Arpizio (D'Arpizio, 2008) defines the 3A's of luxury (Absolute, Aspirational, and Accessible) which constitutes sub-segments of the whole luxury market. In the Absolute luxury, the products must be elitist, represent an icon, and include characteristics of tradition and uniqueness. In the Aspirational luxury, the main value for the customer is represented by aspiration, recognizability, and diversification. The Accessible luxury is characterized by a higher economic accessibility, where the sense of membership and status symbol plays a relevant role.

Within the luxury apparel fashion markets, style and design excellence are achieved by keeping the design activities in-house while collaborating with experienced external designers and stylists (Brun, 2008). The process of New Product Development (NPD) is characterized by a large number of product revisions, in which the continuous interactions between designers, stylists and marketing functions take a long time and iterations (Sen, 2008). Revisions and modifications are realized also when the final product is already in the market, in order to align the production to the customer demand (e.g. change colors for a model).

NPD is characterized by high seasonal demand, which depends on the seasonal nature of fashion products. The whole NPD process is managed at least four times per year - one time for each season - with a short TTM (15 months in the apparel industry, 12 months in the leather industry)

In the luxury fashion market, NPD is a comprehensive process which includes the following five stages:

i. Design
ii. Modeling/prototyping (for realizing the demonstration products to be shown at the fashion fairs)
iii. Detailed engineering: Many companies complete the engineering phase at the same time as the sourcing phase. For example, in most cases, the generation of the final Bill of Material (BOM) takes place when orders have already been launched. Once the company confirms the order decisions, the BOM and the raw material purchase orders must be completed quickly. Moreover, the production quantities and decisions may continue to change very rapidly during the fashion shows seasons (i.e. the appearance of a Hollywood star with a particular bag can completely change the prevision regarding the sold units of that product, and this can happen also the last day of the last fashion fair). In some companies, the engineering phase is completed for all the products before the beginning of the fashion fairs. Although this strategy permits quick management of the sourcing activities, it likely overloads the engineering staff with less valued activities, as the actual goods produced are usually a small percentage of the all items presented at the fashion shows.
iv. Material sourcing: The sourcing phase is very peculiar, as its duration can fluctuate from 2 weeks up to one and a half months depending on the duration of the commercial launch, which generally takes place in conjunction with the fashion shows and fairs (e.g. Pitti Florence fair, Milan and Paris fashion weeks, etc.). At the beginning of the sourcing phase, a provisional and generic order of raw material is submitted to the suppliers. The confirmation of the raw material quantity is given at the end of this phase, with a 20-30% maximum difference in quantity from the provisional order to the final order. During this very short period, the company board receives information on the number of sold units for the current season and must quickly decide which products will be produced for the upcoming season, thus defining the raw material order needs.
v. Production and distribution: The production phase usually lasts 3-4 months, beginning when material sourcing is completed.

PRODUCT LIFECYCLE MANAGEMENT IN THE FASHION CHAIN

The Italian Luxury Leather Industry

"Made in Italy" is synonymous with quality, fashion and luxury. In previous years, Italian companies (most of them Small and Medium Enterprises, SMEs) of the fashion market have concentrated their activities in the highest value added luxury segments, guaranteeing the success of the national industry. Still today, despite the global crisis, luxury is still supporting the economic development of Italy (Brun et al, 2008), (AIMPES, 2009).

In the luxury fashion segment, the leather industry is one of the most active compartments. Pushed by an increasing global demand, the first decade of the Twenty-First century was a time for many fashion brands to introduce leather products in their product lines. Leather treatment has a long history in Italy, particularly in the region of Tuscany where many craft laboratories have existed for more than one century. The Italian leather industry is composed of shoe-making, suitcase and baggage making, and tanning treatment. The luxury leather market may be used as a reference market for the entire luxury fashion market, as addressed below.

This research has been conducted through an empirical, explorative research. It was carried out through personal interviews which were supported by a reference questionnaire. The reference questionnaire, adopted as an interview guideline, is a mix of open and closed-ended questions, composed of the following three sections:

i. General data about the company and the market.
ii. Details about CAx and CPDM tools adopted in the NPD process.
iii. Information about the deployment of PLM (CAx and CPDM) projects.

In this way, the data has been explicated through the analysis of the level of application of PLM, in terms of CAx and CPDM installed solutions, with the tentative purpose to make a comparison between unconventional PLM-oriented industries and more traditional PLM-oriented contexts.

The research has been limited to the luxury leather sector, with companies producing only finished leather goods (i.e. bags, suitcases, shoes, wallets and belts). According to the national leather association, AIMPES, (Associazione Nazionale Pellettieri) this sector is composed of 54 firms in Italy, including transnational companies belonging to the luxury sector with at least one business unit and at least one production plant in the country. All these companies are very well know international luxury names such as Burberry, Roberto Cavalli, Dolce & Gabbana, Dior, Fendi, Ferragamo, Etro, Furla, Gucci, Missoni, Montblanc, Prada, Sergio Rossi, Tod's, Trussardi, Versace and Ermenegildo Zegna - just to name a few. All of the 54 companies have been contacted and 20 of them (37% response rate) agreed to be interviewed and anonymously surveyed. Interviews have been conducted with product managers of the leather business units and/or to the responsible party of the relative technical departments. In some cases, the questionnaire has been submitted also to the responsible party of the ICT departments, when existing (less than the 25% in the sample). The data of the survey have been analyzed through qualitative instruments, hereafter presented and discussed.

All the surveyed companies operate in the luxury leather market, at the top of the fashion pyramid (Couture). According to the classification (D'Arpizio, 2008), 60% of the sample operates in the Absolute luxury segment and the remaining 40% in the Aspirational segment, and they serve global markets, typically within single brand shops.

The sector is composed of international companies, most of which are fully integrated and multi-product, belonging to multi-brand groups, with considerable amounts of capital available. Using the definition of value chain of (Saviolo and Testa, 2005), 10 companies are well known Integrated Griffe, 6 are part of Industrial Integrated Groups and 4 are Pure Brand Owners. Most of them (70%) are Italian owned companies, while 30% are business units of international transnational companies (25% based in Europe, 5% based in the US). In all the cases, the leather segment is just one of the luxury segments served by the companies. 90% of the sample companies are located in central Italy's region of Tuscany, while the remaining 10% are located in Northern Italy.

In terms of products, 70% of the sample produces only leather bags and suitcases, while the remaining 30% make leather shoes too. The majority of the sampled companies (55%) identify their core business in the leather sector (35% leather accessories, 20% leather footwear), while the remaining 45% companies report that clothing is their core business, and leather is just a slice of it.

An interesting "volume" parameter is the number of different Stock Keeping Units (SKUs) processed during the fashion season by the companies. A SKU is a single product code realized and sold during a fashion season (a "collection") and encompasses all the possible factors, such as color. For example, if the same model of a bag is produced blue and brown, then two different SKUs might be defined, one for the blue bag and one for the brown bag. More than 50% of the companies realize collections of more than 500 SKUs per season. Among them, 20% of the sampled companies define more than 2,000 SKUs per season. Only 3 companies show fewer than 100 SKUs per season. In the new SKUs per season, some of them are directly related to SKUs of the previous season (e.g. just a new color – black – for an old model bag) and they are defined as Carry Over (CO) units - items produced continuously every year. More than 70% of the companies have a reduced CO (less than 30% on the total SKU per season), highlighting the constant innovation of the companies to continuously look for new fashion trends-waves for satisfying their customers.

It is widely recognized (Brun et al, 2008) that the fashion sector widely applies outsourcing strategies (Table 2). In more than 90% of the sample, production is partly outsourced to external production laboratories, generally located near the main plant and primarily based in Tuscany. On average, a company has more than 35 external small and medium laboratories working in its supply chain. Moreover, almost 50% of the companies design their products with the support of external designers.

Table 2. Percentage of outsourced activities

Outsourced activities	%
Production	95%
Prototyping	70%
Product development	50%
After sales	30%
Logistic	30%

ICT in the Leather Industry

In terms of ICT tools, all the companies have an ERP (Enterprise Resource Planning) suite installed in order to manage their actives and passive order cycle. In the NPD process, a significant difference has been observed between CAx and CPDM tools, for instance, 80% normally apply CAx tools (80% CAD, 45% also CAM), while only 40% of them have installed a CPDM solution.

45 percent of the companies adopt a CAD 2D, 30% adopt both CAD 2D and 3D, while 5% apply only a CAD 3D, and 20% of the cases do not have a CAx suite in their design departments. These companies just agree on the concept of the product and then totally outsource the detailed design and engineering to external professional designers. The leather sector calls for a CAD tool specifically designed for the leather/textile activities. Fewer than 20 percent of the sampled companies apply a multipurpose CAD (i.e. SolidWorks by Dassault Systèmes, ThinkDesign by Think3), while the majority of the sample uses a list of dedicated tools, most of them created in Italy (i.e. Mozart CAD, Romans CAD, Naxos, AIMPES CAD, Bitron CAD, etc.). Generally, when it exists, the installed CAM is provided by the same vendor of the CAD system. Furthermore, 40 percent of the companies have installed a CPDM solution.

Collaborative platforms enable cooperation among functions and business processes, generally proving a common data warehouse (vault) for sharing data, metadata and information. CPDM platforms can be implemented with different degrees of functionalities, supporting knowledge sharing in unstructured (e.g. simple file sharing email-based) or structured ways (e.g. automatic workflow). Further, CPDM platforms are used within small working groups (e.g. only the technical department) or larger teams (e.g. involving key actors of other phases, like marketing, production, etc.). In the analysed sample, 5 companies installed a CPDM solution (a traditional PDM platform) in the technical department, just for storing and sharing product data (i.e. drawing and 3D models) and managing NPD projects. In fewer than 5% of the companies, commercial web-based CPDM platforms have been installed for enabling collaboration along the NPD process, supporting product data (e.g. models and drawings) and metadata (e.g. BOM and configuration rules) sharing among the different phases (design, sourcing and production). In most of the cases, a deep customisation of CPDM commercial platforms was needed to adapt the available solution to the peculiarity of the industrial context.

Most of the benefits reported by the companies regarding CAD application are the reduction of TTM and process quality improvements. The former is given more appreciation, since it enables faster realization of models and easier re-use of produced parts. The concept of process quality has been explained by the interviewees as better management, better storage of models, more security, and increased control on stage. Regarding CAD 3D, it was evident that, in this sector, solid modelers are used mostly for marketing purpose (i.e. display of 3D models that could replace, in some cases, the real prototype), while CAD 2D is still enough for the objectives of the designers. The main registered benefits related to the implementation of CAM application are time reduction (cutting speed), quality improvement (more precision and less errors) and costs reduction (reduced waste materials and use of unskilled manpower). In summary, CAD and CAM tools contribute to increase the efficiency in the NPD process.

With regard to problems encountered by the companies during the introduction process of a PLM system, the resistance of people to change is reported as the most critical one - as stated by more than 50% of the cases. This evidence confirms that, in most cases, the deployment of the PLM approach in a company (in particular in the CPDM segment) is linked to a process or organizational innovation, as happens in other more PLM-oriented sectors. This behavior is often explained in terms of "political resistance", apprehension of managerial control, evading personal initiatives and so on. Other criticisms indicated by companies' managers are related to technical issues and software specifications, even if none of these problems were considered as reasons for any particular project failure.

Pertaining to difficulties met by the companies after the introduction of a PLM tool, certain problems existed. In half of the analyzed projects, the introduction of the system do not reduced the employee's resistance, which felt their benefits during the start-up of the projects. In more than half of the projects technical problems have been detected, during the introduction's phase, because PLM tools needed users dexterity and involves high personalization and integration (in particular CPDM platforms) in this industry. At the end of the project, most of the surveyed companies were satisfied by the introduction of the new software products. Only 10% of the companies reported the PLM introduction project as a failure.

Two cases of failures are analyzed in this research, both related to the introduction of basic CPDM platform. In the first case, the causes were related to technical issues (the proposed solution did not meet company's needs), while in the second case the resistance of the business units and less of a commitment of the company managements have to be considered as the main reasons for the projects failure. These two cases confirm a common thinking widely debated in literature and common to ICT projects: during the introduction of ICT tools, human elements are the critical factors in determining the success or the failure of the projects. Furthermore, a good management and a strong commitment are needed to guarantee the success of implemented projects.

Software specification is another crucial aspect to ponder. In the first part of a project, an important step is the definition of how a commercial tool has to be personalized in order to full fill a company's particular needs. In many cases, the effort needed to support this part of a project was underestimated by managers, causing problems and delays.

Other causes reported by the companies as possible criticalities in the introduction of CPDM systems often had to do with the dimension of a company. One common thought is that, if the collections are made of fewer SKUs, or a division has fewer personnel, it does not seem relevant to introduce this kind of collaborative systems. On the other hand, the commitment coming for a corporate strategy, even if not supported by real company's needs, is more effective for adopting CPDM software, rather than for CAx tools. In fact, CPDM systems enable collaboration among departments, functions and sites. In a small context normal communication could be enough to foster collaboration, while in big contexts (like corporations and transnational companies) collaboration might be pushed and physically supported by using dedicated ICT CPDM platforms. CAx tools are directly related to the activities which they are designed for, so their adoption is independent by a company structure and dimension.

Despite the criticalities mentioned above, CPDM systems can offer many benefits to companies. CPDM platforms generally permit a better cost control of a product, beginning from the design phase, and provide the opportunity to increase cross-functional collaboration and establish a better control on internal processes. CPDM systems guarantee to the user that he/she can assess a structured and shared archive where he/she can easily find the right product information (i.e. the right version of the SKU), thus enabling a more efficient use of time, a reduction of the risks and errors in the management of the BOM. A typical benefit achieved with a CPDM system in traditional PLM-oriented sectors is a general improvement in the component and part reuse and management. In this particular industry, because of the peculiarities of the fashion sector itself, in which market demands for new products, forms, and materials, such kind of benefit was not considered useful by the companies.

ANALYSIS AND DISCUSSION

Within the sample, those companies with the most advanced and integrated CAx solutions (i.e. CAD 2D and 3D integrated and also directly connected with CAM) are managing more SKUs per year, with more products are being designed continuously using sophisticated ICT solutions.

In order to evaluate and classify the attitude of a company in the adoption of PLM tools, a proxy variable has been created and named as ICT Maturity Degree. This variable has been defined considering the interaction between the degree of functional coverage and the existing integration among the software of the company, where functional coverage is a parameter linked to the internal processes supported by ICT tools, and integration means the system capability to share information and to communicate inwards and outwards the company. Regarding the integration, the levels have been defined as:

- None: companies without any CAx and CPDM systems.
- Low: in case of isolated CAx tools and/or CPDM systems.
- Medium: when CAx tools and/or CPDM systems are partially integrated among themselves or with other ICT suites.
- High: in case of fully integrated ICT solutions.

In a similar way, four levels of functional coverage, with increasing complexity, have been defined as:

- A zero level, in which companies without any CAx and CPDM systems have been included and classified with "Inexistent" functional coverage.
- A first level (Low), in which takes place a simple transaction of drawings and data sheets among business functions, implemented with an extemporaneous use of e-

mail or data warehouse. Companies without a CPDM system belong to this level.

- A second level (Medium), where there is a more complex management of data and metadata among company's functions; at this level is required the existence of a relational database and a data warehouse system, able to manage controlled and restricted access, without using automated workflow. Companies with traditional commercial PDM belong to this group.
- A third level (High), with an intensive use of automated workflow among functions, data warehouse systems able to manage of product data and other complex information. In the present study, this level represents the highest definition of PLM-oriented ICT system.

Crossing functional coverage and integration (Table 3), it is possible to define four levels of ICT Maturity Degree, which are:

- Group 1(None): in case where both functional coverage and integration are absent.
- Group 2 (Low): with a low level of functional coverage associated to low and/or medium integration level.
- Group 3 (Medium): when both parameters are medium.
- Group 4 (High): in the following cases: high functional coverage and a medium and/or high integration level; medium functional coverage and high integration level.

The 40 percent of the sample have a low maturity degree, generally due to a limited use of collaborative software. The remaining 60% of the sample is equally divided among the other three levels. The highest level corresponds to the application of a comprehensive PLM approach.

Some of the most interesting results of the survey come from the comparison of the ICT Maturity Degree variable with other companies' parameters, as the number of SKU and the percentage of CO. Table 4 and Figure 1 map the ICT Maturity Degree with the number of SKU, in Figure 1 the bullets represent the companies. Even if the nature of the research cannot demonstrate this sentence, a direct relation between the number of SKUs (then the dimension of the seasonal collection) and the ICT Maturity Degree seem to exist. That is, the bigger is the collection, a more sophisticated and integrated ICT infrastructure is needed. In Figure 1 there are two outliers (i.e. two companies with high SKU, but without a CPDM system). The empirical nature of the research does not explain this phenomenon, but the possible causes are the budget constraints and a general mistrust of ICT solutions which still exist in some companies of the fashion market.

Table 5 crosses ICT Maturity Degree with CO percentage. From these results, two possible relationships can be supposed:

Table 3. Analysis of the ICT Maturity Degree

Level of functional coverage	**Inexistent**	**I (Low)**	**II (Medium)**	**III (High)**
Level of Integration				
None	20%	-	-	-
Low	-	25%	-	-
Medium	-	15%	20%	5%
High	-	-	5%	10%

Table 4. SKU classes and ICT Maturity Degree

ICT Maturity Degree	Group 1	Group 2	Group 3	Group 4
SKU classes				
< 100	25,0%	25,0%	-	-
100 - 200	25,0%	12,5%	-	-
200 - 500	25,0%	25,0%	25,0%	-
500 - 1000	-	25,0%	50,0%	-
1000 - 2000	25,0%	12,5%	25,0%	-
> 2000	-	-	-	100%

- A small CO percentage could indicate high variability in the collection, which calls for more sophisticated ICT tools.
- A high CO percentage could indicate a relevant continuity among collections, which could call for sophisticated CAx for enabling a better design reuse.

As mentioned, NPD process (or part of it) is often outsourced in the leather industry. Table 6 maps ICT Maturity Degree with NPD outsourcing. Obviously, 100 percent of the companies of Group 1 operate in outsourcing, even if in many other cases (with a higher ICT Maturity Degree) part of the NPD process is realized externally. Evidently, PLM (CAx and CPDM) systems enable for a better management of the relationship with suppliers (e.g. for CAD models display, collaborative management of technical data, etc). However, it might be said that in most of the sample cases unstructured traditional web-based technologies (like e-mail and FTP) are used to exchange product data with outsourcers, while structured solutions (e.g. CPDM workflows) are not so well diffused.

Analyzing the relationship between the size of the leather business unit – measured as the number of people belonging to the business unit and ICT Maturity Degree, it appears that all the cases without any ICT tool, and 87.5% of those who have only CAx tools, have a staff under 100 employees (Table 7). Moreover, in all the cases with CPDM systems installed have more than 100 people. These cases can be further divided into two groups, one with basic CPDM systems (66.7% of the companies), and the other with advanced CPDM solutions (100% of the companies). The data seem to show a link between the size of the business unit and the presence of PLM solutions despite their level of complexity. The same results have been reached by evaluating through the number of people dedicated to the NPD process. The implementation of CPDM systems seems to be linked also with the dimensions of the design and technical department (Table 8). Having more

Figure 1. SKU versus ICT Maturity Degree

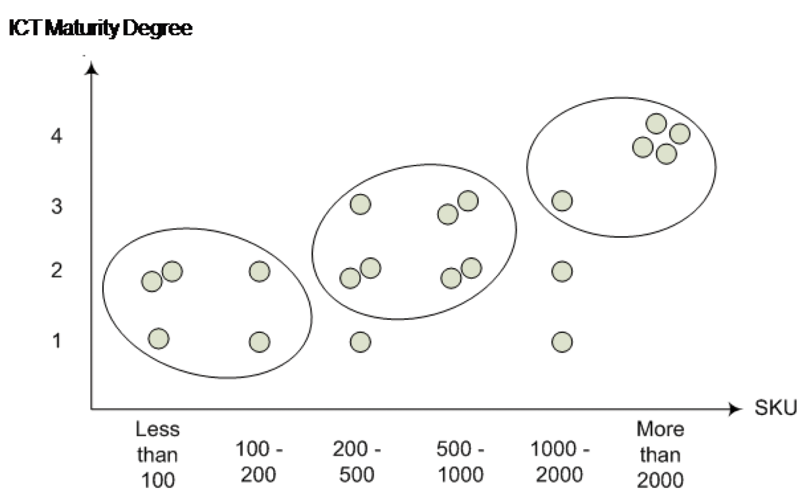


Table 5. CO percentage versus ICT Maturity Degree

ICT Maturity Degree	Group 1	Group 2	Group 3	Group 4
CO percentage				
0-15%	-	25,0%	16,7%	50,0%
15-30%	50,0%	37,5%	83,3%	-
30-45%	-	-	-	50,0%
> 45%	50,0%	37,5%	-	-

Table 6. Outsourcing of the NPD process versus ICT Maturity Degree

ICT Maturity Degree	Group 1	Group 2	Group 3	Group 4
Outsourcing NPD				
NO YES	- 100%	63% 38%	50% 50%	75% 25%

designers and engineers involved in the leather business unit, the higher the ICT Maturity Degree appears. Due to the explorative nature of the survey, however it is not possible to confirm this hypothesis, even if many indicators seem to confirm this thesis. Such a relation (i.e. company dimensions versus ICT Maturity Degree) represents one of the biggest limits of PLM in this sector. The available PLM technologies seem to be still dedicated to big companies, while SMEs are excluded from this market.

Others possible correlations have been sought e.g. the core business of the brand and the ICT Maturity Degree, but any relevant connection regarding adopted tools has come out. It appears that 100% of the companies without any tools have clothing as their core business: this evidence could partially explain small ICT investments in the leather sector, but it doesn't provide any information about a possible higher ICT Maturity Degree of the core division.

One of the objectives of the research was the identification of possible trends in adopting CAx systems and CPDM solutions in fashion sectors. In this way, during the research, past and future ICT projects regarding the NPD process have been also investigated. Regarding past projects (Table 9), results show that the last tool introduced by the companies belonging to Group 2, having only CAx tools, in 75% is the CAD system itself,

Table 7. Leather Business Unit Personnel versus ICT Maturity Degree

ICT Maturity Degree	Group 1	Group 2	Group 3	Group 4	Total Companies
Leather BU Personnel					
< 50 50-100 101- 250 251- 500 > 500	100,0% - - - -	62,5% 25,0% 12,5% - -	16,7% 16,7% 33,3% 16,7% 16,7%	- - - 50,0% 50,0%	50% 15% 15% 10% 10%

Table 8. Number of Business Unit Designers versus ICT Maturity Degree

ICT Maturity Degree	Group 1	Group 2	Group 3	Group 4	Total Companies
Number of BU Designers					
< 50 50-100 100-250 250-500 > 500	100% - - - -	62,50% 25,0% 12,5% - -	- 16,7% 33,3% 16,7% 16,7%	- - - 50,0% 50,0%	50% 15% 15% 10% 10%

percentage equally divided among 2D and 3D. Companies owning basic CPDM solutions, in 25% of the cases have recently introduced 2D CAD tools, as substitutions of obsolete software or as integrated system to CAM (system implemented in 25% of these companies). The remaining 25% have recently adopted their current CPDM solutions. The results of Group 4 is interesting too: in this case, for 75% of the companies, the last implemented system is the CPDM solution, while in the remaining part, the last tool is a CAD 3D, as a renewal or extension of an existing structure.

These results show that, in the luxury leather sector, the introduction of collaborative solutions is recent and currently in progress, while CAD systems have been adopted since several years in most of the cases. For that matter, CAD 2D tools are still sufficient for the industrial needs and substitution with CAD 3D seems to be still far away.

A specific set of questions were related with the investigation of ICT projects planned by the companies in the next 3 years, the results are interesting and can be summarized in the following statements:

- 75 percent of companies without any tools in their NPD are not planning to introduce them in the next years. This evidence can be justified in some cases with the presence of other ICT projects currently in progress, or due to budget and/or human resources limitation for a new project, or because they are considered not strategic for the firm by the management. The remaining 25% have planned the introduction of basic CPDM systems.
- All the companies with only CAx tools are preparing new projects. Where in 30% of the cases extension of CAD or CAM tools are considered in order to cover additional functionalities. While in the remaining 70%, they consist of projects for the introduction of CPDM solutions. In the end, it is important to underline that the cost of these projects, due to their extension, is always considered significant by the management of the companies.
- Companies belonging to the third group, which own a CPDM system, 75% of them have projects regarding these tools. These projects are mainly related to the extension of functionalities or to improve the integration level, and in most of cases changing their own basic solution into a most advanced one.
- Focusing on companies with advanced CPDM solutions (only 3), despite the common thinking, results that all of them have already planned projects of functional coverage extension or improvement of the integration level. In authors' opinion, this evidence confirms that the benefits obtained by the companies are the main incentive to promote a full implementation of the PLM business model. Obviously these kinds of projects, such as those of the previous paragraph, consisting of extensions of solutions already implemented, will be low-

Table 9. Last ICT tool installed by the company (within the four groups of ICT Maturity Degree)

ICT Maturity Degree	Last ICT tool installed						
	ERP	CAD 2D	CAD 3D	CAM	Basic CPDM	Advanced CPDM	Other
1	100%	-	-	-	-	-	-
2	-	37,5%	37,5%	-	-	-	25%
3	25%	25%	-	25%	25%	-	-
4	-	-	25%	-	-	75%	-

er than the size of those introduced from scratch.

Moreover, in most professional literature, dominated by vendors and consultants publications, there is a recurrent interest on the application of PLM (in particular in its CPDM segment) in fashion sectors. Some contributions ((e.g. (Barret, 2007), (Hojlo and Suleski, 2008), (Suleski et al, 2007)) state that fashion will be the next PLM market. This exploratory research partly confirms such an interest, though PLM is at the fledging stage in these sectors. However, it might be noticed how in the last five years many companies in the luxury leather industry have changed their CAx tools and have introduced their first CPDM systems, and many others plan to introduce a CPDM solution in the next years. This means that PLM is certainly gaining growing interest across these companies.

CONCLUSION

In this contribution the main results of an empirical research concerning the application of PLM solutions for supporting the NPD process in the luxury leather sector are reported. Even if the luxury leather sector is a niche market of the largest fashion sector, it can be considered well representative of the criticalities and features of all the other sectors. Despite this, obviously, the analysis of such a specific market is one of the most relevant limits of the research, which has an explorative purpose and can't be considered as exhaustive.

However, in its limitations the research underlined the main peculiarities of PLM projects in the fashion industry, identifying some commonalities and differences that characterized this sector in comparison with more PLM-oriented sectors, like the mechanical one.

From this preliminary analysis some interesting new research items have been identified. For example, it is interesting to notice how after sale service is not considered at all in the ICT strategies of the analysed companies. A common thought widespread among the interviewees is that the non-durable nature of products implies a short lifecycle. This evidence moves companies to have "lack of memory" on themselves. Thus after sales, in spite of other sectors where maintenance is a strategic phase of the lifecycle, in the luxury fashion sector is currently a marginal phase of the process, where no resources focus on it. In this sector, after sales is still considered as a service that the companies as to provide, and not as a source of further earnings or competitive benefits (e.g. aerospace). The adoption of a "complete PLM approach" is still too far to be implemented in fashion contexts as in more PLM-oriented sectors such as aerospace (Tang et al, 2008) and automotive (Alemanni et al, 2008).

Finally, it is evident that the fashion context has a particular NPD process, which urges for a specific adaptation of PLM solutions. Such a peculiarity needs a new reference model different from the traditional mechanical NPD process. It will thus be interesting to conduct more in-depth analysis on the topic. For instance, more samples can be collected and rigorous statistical analysis can be conducted to reveal more important academic and managerial insights. Moreover, as motivated by the findings of this exploratory research, mathematical models can be constructed for further studies.

REFERENCES

Alemanni, M., Grimaldi, A., Tornincasa, S., & Vezzetti, E. (2008). Key performance indicators for PLM benefits evaluation: The Alcatel Alenia Space case study. *Computers in Industry*, *59*(2-3), 833–841. doi:10.1016/j.compind.2008.06.003

Associazione Italiana Pellettieri. (2009). *Nota Congiunturale* 2009. Retrieved from www.aimpes.it (in Italian)

Barrett, M., Suleski, J., & Burkett, M. (2007). *PLM for apparel, footwear, and soft goods: Meeting the demands of new product introduction*. Retrieved from www.amrresearch.com

Bruce, M., Daly, L., & Towers, N. (2004). Lean or agile? A solution for supply chain management in the textile and clothing industry. *International Journal of Operations & Production Management*, *24*(2), 151–170. doi:10.1108/01443570410514867

Brun, A., Caniato, F., Caridi, M., Castelli, C., Miragliotta, G., & Ronchi, S. (2008). Logistics and supply chain management in luxury fashion retail: Empirical investigation of Italian firms. *International Journal of Production Economics*, *114*, 554–570. doi:10.1016/j.ijpe.2008.02.003

Burresi, A., & Ranfagni, S. (January 2006). *Modelli di supply chain e reti cognitive per l'innovazione di prodotto nel sistema moda di fronte alle logiche emergenti di decentramento produttivo – Le tendenze del marketing in Europa*. Paper presented at the meeting of the International Congress Marketing Trends, Venice, Italy, (in Italian)

Caniato, F., Caridi, M., Castelli, C., Golini, R., Spina, G., & Sianesi, A. (April 2007). *Supply chain strategy in the luxury industry: Do consolidated models fit?* Paper presented at the meeting of the 16th International Purchasing and Supply Education and Research Association (IPSERA) Conference, Bath, UK.

CIMdata. (December 2008). *PLM future trends*. Paper presented at the meeting of DS Israel Annual Conference, Herzliya. Israel. Retrieved from www.CIMdata.com

D'Arpizio, C. (2008). *Altagamma e Bain & Company worldwide markets monitor*. Retrieved from www.altagamma.it (in Italian)

De Toni, A., & Meneghetti, A. (2000). The production planning process for a network of firms in the textile-apparel industry. *International Journal of Production Economics*, *65*(1), 17–32. doi:10.1016/S0925-5273(99)00087-0

Fernie, J., Moore, C. M., Lawrie, A., & Hallsworth, A. (1997). The internationalization of the high fashion brand: The case of Central London. *Journal of Product and Brand Management*, *6*(3), 151–162. doi:10.1108/10610429710175673

Gereffi, G., & Memedovic, O. (2003). *The global apparel value chain: What prospects for upgrading by developing countries*. United Nations Industrial Development Organization. Retrieved from www.unido.org

Ghemewat, P., & Nueno, L. J. (2003). *Zara: Fast fashion*, (pp. 1-35). Harvard Business School Case (9-703-497).

Grieves, M. (2006). *Product lifecycle management – Driving the next generation of lean thinking*. USA: McGraw-Hill.

Hojlo, J., & Suleski, J. (2008). *PLM Spending in apparel: Striving to achieve value chain orchestration*. Retrieved from www.amrresearch.com

La Marca, P., & Palamara, G. (2005). *Strategie di nicchia nel settore moda. Quaderno di ricerca n. 4.* Dipartimento di Ricerche Aziendali dell'Università degli Studi di Pavia. Retrieved from http://economia.unipv.it/ ric-az/ (in Italian)

Ma, Y., & Fuh, J. Y. H. (2008). Product lifecycle modelling, analysis and management editorial. *Computers in Industry*, *59*(2-3), 107–109. doi:10.1016/j.compind.2007.06.005

Moore, C. M., & Birtwistle, G. (2004). The Burberry business model: Creating an international luxury fashion brand. *International Journal of Retail and Distribution Management*, *32*(8), 412–422. doi:10.1108/09590550410546232

Nueno, J. L., & Quelch, J. A. (1998). The mass marketing of luxury. *Business Horizons*, (November–December): 61–68. doi:10.1016/S0007-6813(98)90023-4

Richardson, J. (1996). Vertical integration and rapid response in fashion apparel. *Organization Science*, *7*(4), 400–412. doi:10.1287/orsc.7.4.400

Saaskvuori, A., & Immonem, A. (2005). *Product lifecycle management*. Berlin/Heidelberg, Germany: Springer.

Sapuan, S. M., Osman, M. R., & Nukman, Y. (2006). State of the art of the concurrent engineering technique in the automotive industry. *Journal of Engineering Design*, *17*(2), 143–157. doi:10.1080/09544820500275511

Saviolo, S., & Testa, S. (2005). *Le imprese del sistema moda - Il management al servizio della creatività*. Bologna: Etas Libri. (in Italian)

Sen, A. (2008). The US fashion industry: A supply chain review. *International Journal of Production Economics*, *114*, 571–593. doi:10.1016/j.ijpe.2007.05.022

Silverstein, M. J., & Fiske, N. (2003). *Trading up: The new American luxury*. New York, NY: Portfolio/Penguin Group.

Stark, J. (2005). *Product lifecycle management – 21st century paradigm for product realisation*. London, UK: Springer-Verlag.

Suleski, J., Hojlo, J., & Burkett, M. (2007). *PLM technology adoption in apparel: Turbo-charging investments for fast fashion*. Retrieved from www.amrresearch.com

Tang, D., & Qian, X. (2008). Product lifecycle management for automotive development focusing on supplier integration. *Computers in Industry*, *59*(2-3), 288–295. doi:10.1016/j.compind.2007.07.002

Vona, R. (2004). Modelli di management e dinamiche competitive nel sistema moda: Il caso Zara. *Finanza Marketing e Produzione, 2*. (in Italian)

This work was previously published in Fashion Supply Chain Management: Industry and Business Analysis, edited by Tsan-Ming Choi, pp. 270-285, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 76
Knowledge Dissemination in Portals

Steven Woods
Boeing Phantom Works, USA

Stephen Poteet
Boeing Phantom Works, USA

Anne Kao
Boeing Phantom Works, USA

Lesley Quach
Boeing Phantom Works, USA

Category: Processes of Knowledge Management

INTRODUCTION

While there are many aspects to managing corporate knowledge, one key issue is how to disseminate corporate documents with appropriate context. Upon finding an article on a certain subject, for example the material properties of titanium, a reader is likely to be interested in related articles such as applications of titanium or manufacturing methods for titanium parts. Each related article has the potential to increase the reader's knowledge of the subject. Therefore, organizing documents into categories of interest plays an essential role in discovering and interpreting information. Furthermore, categories can be expected to provide historical context, describing how titanium was used in early designs or initial practices used for the repair of titanium parts.

While most large companies make a practice of cataloging and controlling well-established documents, there is a vast set of *explicit information* that has not traditionally been effectively disseminated. This class of information is less formal and may be exchanged, updated, and otherwise managed at the local level. Such information is usually not controlled at the corporate level or governed by the same organizations established to handle more stable information. Processes to disseminate such information tend to be ad hoc or nonexistent. In this article, we discuss the elements necessary to effectively disseminate informal and explicit

DOI: 10.4018/978-1-4666-1945-6.ch076

information not controlled at the enterprise level. While the main emphasis of the article is to promote a general process for the dissemination of this type of material in large corporations, we will use a specific implementation of this process at the Boeing Company as an illustrative example.

BACKGROUND

Traditionally, the dissemination of corporate knowledge has taken a number of different forms. First, there are the methods of classic library science often as implemented by a formal corporate library staffed by trained librarians (Taylor, 2000). This is used for things that are well established: e.g. textbooks, established how-to knowledge on a subject, published papers on a subject, and so on. Second, it has long been necessary to disseminate official policy and procedure through "Command and Control" processes and associated media. In addition, certain industries also require configuration control processes for special classes of information such as product data, drawings, and manufacturing rejection and acceptance documentation. These are all subject to an authentication process, flowing top-down to intended users. A third, extremely important, approach to knowledge maintenance and dissemination has been through mentoring and establishment of departments aligned to technical specialties and communities of interest. This last type of approach is particularly well suited for *tacit knowledge*. A fourth category of knowledge sharing applies to the communication of explicit knowledge among peers but also includes dissemination to management and other reference groups. This method applies to information that is less formal and frequently ephemeral. See Table 1.

This fourth method is of an entirely more fluid nature and, in some cases, represents the majority of a corporation's explicit knowledge. While it is appropriate for the enterprise to disseminate formal information using traditional, formal means, there is a need to disseminate less formal information as well. This informal knowledge often includes the most current information within a company and without adequate dissemination corporate decision-making is likely to fall short. In summary, stable and formal information is well handled by existing library or document release systems. Ephemeral, less formal, and generally less controlled content, while important, is currently only shared across the enterprise by a variety of ad hoc means, if at all.

MAIN FOCUS OF THE ARTICLE

This article focuses on how to systematically share this fourth category of informal and uncontrolled knowledge. The ideal for knowledge dissemination is to make sure information of this type can be well integrated into existing formal content,

Table 1. Summarization of four traditional dissemination methods

Methods	Types of Information	Tacit/ Explicit	Scope of Control	Status of Information
1. Corporate Library	Formal: books, research periodicals, white papers	Explicit	Enterprise-wide	Stable
2. Command and Control	Formal: documentation of standards, regulations, policies, procedures	Explicit	Enterprise-wide	Relatively stable
3. Undocumented	Informal: skill, experience, expertise	Tacit	Local	Both Stable and Ephemeral
4. Uncontrolled	Informal: test notes, presentations, lessons learned, emails	Explicit	Local	Both Stable and Ephemeral

taking advantage of the context that has been created over time by librarians and other formal content management systems. To achieve this, it is necessary to organize this knowledge in a way that is consonant with the information categories of multiple existing systems. This is made possible by using an enterprise *ontology* (see below) or some form of controlled system of keywords which can be mapped to existing vocabularies. Portals, and other tools which allow content aggregation and term mapping, enable sharing of this knowledge at a physical level. It provides search and simple navigation across sources, as well as security services to restrict access as needed. A central shared ontology combined with an interactive text classification tool make dissemination of this knowledge possible.

In the matter of assigning documents to categories, we emphasize the importance of involving subject matter experts. Traditionally, this is done by librarians who are trained to catalog (categorize) content. However, in the case when authors are widely distributed throughout a complex corporate enterprise, we suggest that text classification software can be used by these subject matter experts to facilitate broad knowledge dissemination. The challenge is to provide text classification services which can be used to produce high quality results by users who are not trained in library science.

The essential elements of a distributed dissemination scheme for this type of explicit but informal knowledge are: a portal, an ontology, a text classification system, and a publication process. In combination, these four elements allow autonomous subgroups of a corporate entity to interact with common resources and tools to publish their local work in a way that places it within a context comprehensible to an enterprise audience.

Knowledge dissemination, as used here, applies specifically to explicit knowledge captured in documents from many sources. There are a number of frameworks that address the life-cycle of explicit knowledge (Bock, 1997, O'Dell, 1998), but here we will follow the steps outlined by Mack et al. (2001). In this framework, the basic tasks in knowledge work are Capture/Extract, Analyze/Organize, Find, Create/Synthesize, and Distribute/Share. In particular, text classification has direct benefit to the Analyze/Organize and Find stages and portal services will be the basis of the Distribute/Share stage. As discussed here, knowledge dissemination applies to the Analyze/Organize, Find, and Distribute/Share stages.

Portals

A portal is used to collect content from many different sources, resulting in a virtual collection available through a single point of access. This aggregation of content is perhaps the key characteristic of all portal products. In addition, a portal provides some capability for metadata management whereby tags and values can be directly replicated from source documents or harmonized within the virtual collection by mapping them to a centralized schema. In addition, a portal may permit the addition of metadata based on characteristics of the source system or based on the decisions made by the group about how ontology terms will be attached to documents. The documents themselves remain in their source system, maintained, refreshed, or deleted by the groups that own them.

Other kinds of information systems besides portals can be useful for knowledge dissemination. For example, several enterprises make use of Wikipedia-style systems (Burton, 2008, Gardner, 2008) Also, influenced by the success of applications such as MySpace and Facebook on the internet, some enterprises are building these types of social computing applications on their corporate intranet, for example IBM's Beehive (Associated Press, 2008). However, portals, in one form or another, are well suited for publishing a distributed collection based on the intellectual products of many subgroups. Further, because of their flexibility in combining a variety of tools and services, portals can be customized to create a rich knowledge-sharing environment. In addition, they can be extended to support personalization services, which would allow even more focused dissemination.

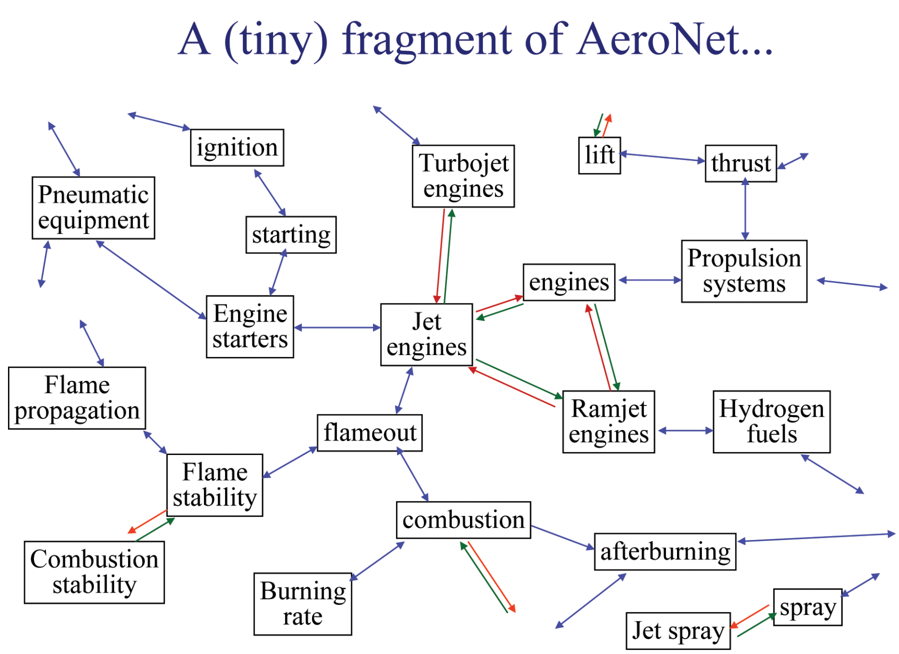
Thus, portals are a natural element to aid in knowledge dissemination. They can be used to achieve the key goal of achieving awareness (Alavi & Leidner, 1999, Prusak, 1997). Indeed, creating awareness is a goal of dissemination and is a prerequisite to collaboration or further synthesis.

Ontologies

To produce an organization of corporate documents that can be readily shared, it is essential to have some broadly shared source of terms that represent the concepts in our shared ontology. Ideally this would come in the form either of a corporate *taxonomy*, a corporate *thesaurus,* or both. What is minimally necessary is simply a list of controlled keywords expressing the topics that are important to the enterprise. We refer to the concepts expressed by these keywords as *knowledge categories*. Ideally, there is some kind of structure or organization to the knowledge categories, for example, a *generalization-specialization hierarchy*. Such a specification of the things that can be talked about and their relationships is referred to in artificial intelligence as an ontology. There is a strong advantage to organizing documents this way, because it enables much easier and better quality search, and greater use of the knowledge in those documents. In the Boeing case, there were about 60,000 knowledge categories, as well as generalization-specialization relationships, synonymy relationships, and more general "is-related-to" relationships, so that the ontology is somewhat richer than a hierarchy (see Figure 1), (Clark et al., 2000).

This approach benefits most when it can build on an existing cataloging scheme, such as one maintained by an enterprise library system or a professional organization (e.g., the Institute of

Figure 1. The AeroNet ontology contains more than 60,000 concepts used as categories for text classification and subsequent knowledge dissemination. This figures shows a small fragment of the total ontology.



Electrical and Electronics Engineers), or a taxonomy provided by an industrial consortium (e.g., the Air Transport Association). The ontology used in the Boeing realization of the approach derives from the corporate thesaurus, a list of controlled keywords sanctioned by the Boeing technical library system, together with the relationships among them. By using a set of controlled keywords created by corporate librarians, not only was an immense amount of time saved in constructing the taxonomy but the system was also able to leverage off of their collective and accumulated expertise.

However, with some additional effort, categories can be established and applied even when such a resource is not available or is not sufficiently complete. In this case, one approach is to use text clustering and automatic summarization (Kao et al., 2004). Text clustering serves to group documents together based on the words they contain. The words with the greatest frequency or strength either throughout or near the center of each cluster can summarize the documents in that cluster. Using this summary, domain experts, knowledge administrators or managers, or librarians can decide which clusters represent important categories to the enterprise, and supply standard names for those categories. If desired, they can further subdivide or group the resulting categories giving generalization-specialization relationships, and shared summary words may suggest other relationships. Of course, while this approach will provide more consistency within the enterprise, it will lose the historical context provided by leveraging off of existing ontologies.

A folksonomy, as created by social tagging applications, is another option as a source of terms when an enterprise has no pre-existing ontology, thesaurus, or taxonomy (Vander Wal, 2007, Hayman, 2007). Vander Wal described this as a "user-created bottom-up categorical structure development with an emergent thesaurus." An argument against using folksonomies as a source of terms for a formal ontology is that folksonomies typically contain tags that are mostly personal (e.g. "to-read"), or subjective (e.g., "hilarious" or "awesome"); however, Al-Khalifa and Davis (2007) and Sen et al. (2006) find that about 2/3 of user-generated tags are content related (e.g., "lean+", or "web2.0") and, therefore, provide a useful source of terms for a corporate ontology. As with the document-based text clusters mentioned above, folksonomies do not necessarily capture historically relevant or traditional categories.

Once an ontology is available, the concepts must be attached to documents and used to retrieve those documents. We do not assume that those who are involved in knowledge dissemination, whether as knowledge provider or consumer, have either the training or perspective of librarians. Nevertheless, by allowing participants to interact with a controlled vocabulary in the context of their subject matter of interest and giving them a means to do so easily, the process can improve dissemination of knowledge by grouping topically similar documents together regardless of their original local vocabulary usage. What is required is a tool to aid the author or searcher in finding the appropriate categories.

Text Classifier

An automatic text classifier takes a piece of text (e.g. a memo, an abstract, or a longer document), and based on its features (typically the words it contains) determines automatically which of a set of predetermined categories should be assigned to it. There are a number of different approaches, as summarized in Sebastiani (2002), including naïve Bayes (McCallum & Nigam, 1998), support vector machines (Joachims, 1998), and logistic regression classifiers (Komarek, 2003).

The first step in creating a classifier is to collect a *training sample* of texts that are already associated with the appropriate knowledge categories. At the Boeing Company, the abstracts for about 500,000 corporate documents in the technical library were used as the training sample.

If the categories are not a pre-existing set created by librarians or other standard professional body but have been developed as described above using a text clustering algorithm, there may still be an easy way to create a training set. If the clustering algorithm generates convex clusters, as does K-means (Hartigan, 1975, Hartigan & Wong, 1979), the central members of the chosen clusters could serve as a training sample for those categories.

A text classifier is then constructed or trained based on the knowledge categories and the training sample of already categorized documents. The text classifier is then made available to both producers and consumers of the documents.

Authors of new documents to be entered into the system can use the text classifier to automatically propose knowledge categories appropriate to their document. In the Boeing implementation of this process, using the Graphical User Interface (GUI) shown in Figures 2 and 3, they first either type in the name and location of their document or browse to enter it into the classifier. The classifier then proposes the most likely knowledge categories ranked in order of likelihood in the box under the "Thesaurus Terms" label (see Figure 2). The Authors can either select these terms or use AeroNet, a semantic network of thesaurus keywords developed at the Boeing Company (Clark et al., 2000), to navigate through the relationships among the categories to find more precise or more general knowledge categories to better describe their document. For example (see Figure 3), the knowledge category "data management" might have been returned by the classifier. By highlighting that and clicking on "EXPAND", AeroNet will bring up specializations like "factory data access system" and "data integration" and generalizations like "management" from which the user can select or use as a starting point for further navigation. A category can be removed from the list by highlighting it and then clicking on the "Remove" button. Finally, clicking on the "OK" button will associate the selected list of categories with the document. An extension of this would be to allow users to have a hybrid system where the users could use tags suggested either from the controlled vocabulary or from a user-generated folksonomy, or enter any word they think would be an appropriate tag if they cannot find anything in the controlled ontology.

During the Find phase, the same text classifier can be used to identify information the user is interested in, without requiring the user to know the controlled vocabulary. Users may search for documents in the Knowledge Management system by entering a natural language query in a similar GUI and then selecting from the knowledge categories suggested to construct a precise query based on the content metadata.

Publication Process

A workgroup needs to have a process that uses the classifier and portal for knowledge to be shared outside the group. Minimally, individuals would be required to add metadata to all documents they wish to publish. They can interact directly with the enterprise ontology where one exists, they may apply user-generated tags from a folksonomy, or they could use the text classifier on a representative sample of their documents to produce an initial down-selection of knowledge categories appropriate for most documents they are likely to produce.

Although social tagging applications are attractive, our main interest is to provide information in context so that the publishing process genuinely constitutes the dissemination of knowledge. By choosing tags that have been used to catalog the broad assets of a corporate library this goal is arguably achieved.

In a discussion of "Library 2.0", (Maness, 2006) suggests that formally controlled terms (e.g., from taxonomy, ontology, or thesaurus) can be profitably combined with terms from a folksonomy and that the approaches are not mutually exclusive. So while folksonomies provide

Figure 2. The Graphical User Interface (GUI) allows users to scan documents and find suggested categories from the AeroNet ontology based on the results of a text classifier.

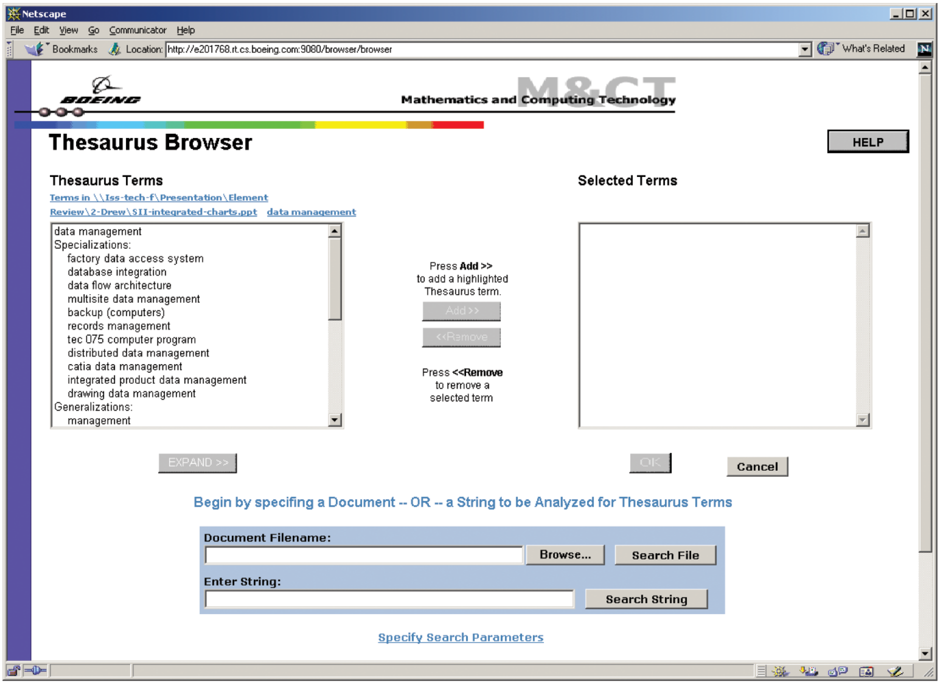
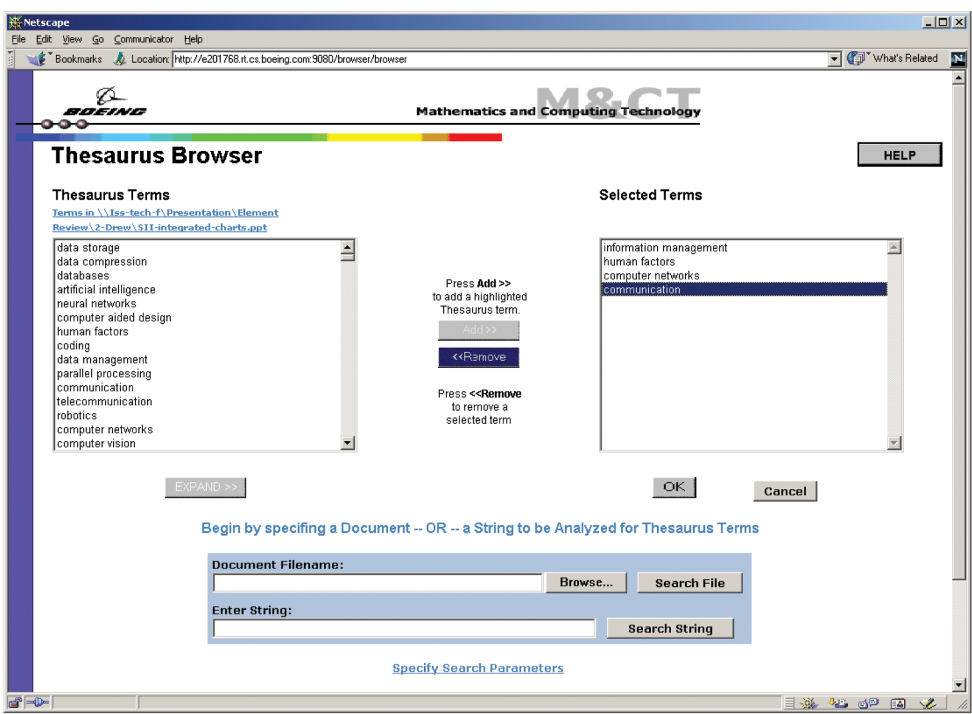


Figure 3. Users can use the GUI to expand terms and then select those that represent the best categories for a given document.



an alternative for enterprises without a library system or a controlled vocabulary, the best case may be where these two practices are combined. In this combined scenario, assets will be placed in the context of historical practice provided by the corporate library but also given meaningful current personal and subjective context through social tagging. Thus context is enhanced in a combined practice. Indeed folksonomies are likely to provide a better representation of current and emerging terminology and this should become an obvious source for updating the authoritative, controlled vocabulary.

FUTURE TRENDS

Effective classification and processes for knowledge dissemination have a positive impact on the evolution of the *semantic web* (Berners-Lee et al., 2001). The basic idea of the semantic web is to make content on the web machine-sensible with regards to its semantics (what the content is about) and with regards to machine reasoning (what the content is related to or implies). In achieving semantics it will be important to classify content in terms of one or more ontologies. Tools involving automatic text classification will probably be essential in making this feasible. Using classification based on existing library practices, as an underpinning, to provide these semantics will have the same advantage mentioned above; the meaning assigned to web content will be more likely consistent with existing catalogs of knowledge. In knowledge dissemination as discussed here, users are able to traverse the relationships in the ontology when choosing terms. However, in the future semantic web, reasoning over one or more ontologies is expected to occur automatically, at runtime, whenever a query is activated. One function of the semantic net will be to improve knowledge dissemination by using software agents to execute queries selectively, taking advantage of both the context of the content and the matching profiles of end users. The different ways that reasoning, and multiple domain ontologies, can be leveraged in the future semantic web should substantially extend what is used in current knowledge classification and dissemination. Nonetheless, the four essential elements discussed here will still be crucial. Without effective text classification, reasoning capabilities can not be expected to provide significant improvements.

CONCLUSION

There is a broad class of explicit but informal knowledge created within almost any enterprise which is not well disseminated by conventional means. Taking advantage of the web infrastructure and the current trend towards social computing, this knowledge can be effectively disseminated using four essential elements: a portal, an ontology, a text classification system, and a publication process. Perhaps most important is a broad, cost effective means to classify content according to either an ontology or folksonomy. Because folksonomies are more likely to contain current, up to date vocabulary and a formal ontology is more likely to reflect historical context, the broadest context can be achieved by combining both as a source of tags. Also, terms within a folksonomy that have garnered substantial consensus, are obvious candidates for subsequent inclusion within the enterprise ontology. We suggest that text classifiers can play a role in either case. Furthermore, while other vehicles such as wikis, blogs, and personal networking applications provide alternatives, portals are well suited to this role as they serve as an information aggregation platform that can combine content from many heterogeneous sources (including Web 2.0 sources), provide role based access, and personalization features. Given this, portals can be used to disseminate knowledge, subject to appropriate publication processes, according to consistent categories. This provides context that links each new item of content to the information previously collected in library catalogs and published from other content resources.

Further, both social tagging and text classification help overcome a substantial obstacle to knowledge dissemination within a large enterprise as users are otherwise reluctant to assign metadata to documents that they author or maintain. By providing aids to classification, basic users can accomplish cataloging tasks without much training, time or effort, and are therefore more likely to do it. In the approach outlined, when based on a formal controlled ontology, the list of proposed knowledge categories is essentially consistent with the practice of professional librarians. Overall, the quality of the user-selected categories is likely to be much better compared to categories selected without assistance from a large ontology. When a folksonomy is used, the knowledge categories will instead favor current usage, rather than a consistent professional practice. Consequently, using these four elements, enterprises have at their disposal a means to disseminate a class of information that is likely to aid corporate decision making and which should qualitatively increase a company's understanding of itself.

REFERENCES

Al-Khalifa, H. S., & Davis, H. C. (2007). Towards better understanding of folksonomic patterns. In Pr*oceedings of the eighteenth conference on Hypertext and hypermedia, Manchester UK* (pp. 163-166).

Alavi, M., & Leidner, D. (1999). Knowledge management systems: issues, challenges and benefits. *Communications of the AIS, 1*(2), Article 7.

Associated Press. (2008). 'Virtual watercooler' makes workplace more fun. Retrieved from http://www.cnn.com/2008/TECH/ptech/05/12/virtual.watercooler.ap/index.html

Berners-Lee, B., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 34–43. doi:10.1038/scientificamerican0501-34

Berry, M., Dumais, S., & O'Brien, G. (1995). Using Linear Algebra for Intelligent Information Retrieval. *SIAM Review*, *37*(4), 573–595. doi:10.1137/1037127

Bock, G. (1997). Knowledge management frameworks. *Patricia Seybold's World Computing Report, 20*(2).

Burton, B., Johnson, E., & Moran, M. (2008). Making It Happen: KM at the Working Level (and Up). *APQC 13th Annual KM Conference*.

Clark, P., Thompson, J., Holmback, H., & Duncan, L. (2000). Exploiting a Thesaurus-Based Semantic net for Knowledge-Based Search. In *Proceedings of the 12th Conference on Innovative Applications of AI (AAAI/IAAI'00), 12* (pp. 988-995).

Gardner, W. D. (2008). Enterprise 2.0: CIA's Secret Intellipedia Has Universal Relevance. *Intelligent Enterprise*, *11*(6), 11.

Hartigan, J. A. (1975). *Clustering Algorithms* (pp. 97–105). New York: Wiley.

Hartigan, J. A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. *Applied Statistics*, *28*, 100–108. doi:10.2307/2346830

Hayman, S. (2007). Folksonomies and Tagging: New developments in social bookmarking. In *Proceedings of Ark Group Conference: Developing and Improving Classification Schemes, 2007, Sydney, Australia*. Retrieved from http://www.educationau.edu.au/jahia/Jahia/home/pid/482

Joachims, T. (1998). Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML-98, 10th Eurpean Conference on Machine Learning, Chemnitz, Germany*, 137-142.

Kao, A., Risch, J., & Poteet, S. (2004). Visualizing Text Data with TRUST and Starlight. *Interface 2004 Conference Proceedings.*

Komarek, P. (2003). Fast Logistic Regression for Data Mining, Text Classification and Link Detection. In *Proceedings of NIPS2003, Conference on Neural Information Processing Systems.*

Mack, R., Ravin, Y., & Byrd, R. J. (2001). Knowledge portals and the emerging digital knowledge workplace. *IBM Systems Journal*, *40*(4), 925–955. doi:10.1147/sj.404.0925

Maness, J. (2006). Library 2.0 Theory: Web 2.0 and Its Implications for Libraries. *Webology, 3*(2). Retrieved from http://www.webology.ir/2006/v3n2/a25.html

McCallum, A., & Nigam, K. (1998). A comparison of event models for naïve bayes text classification. In *AAAI-98 Workshop on Learning for Text Classification*.

O'Dell, C., Grayson, C. J., & Essaides, N. (1998). *If only we knew what we know: the transformation of internal knowledge and best practice*. New York: Free Press.

Prusak, L. (1997). Introduction to knowledge in organizations. In Prusak, L. (Ed.), *Knowledge Organizations*. Boston, MA: Butterworth-Heinemann. doi:10.1016/B978-0-7506-9718-7.50003-2

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*(1), 1–47. doi:10.1145/505282.505283

Sen, S., Lam, S., Rashid, A., Cosley, D., Frankowski, D., Osterhouse, J., et al. Tagging, communities, vocabulary, evolution. In *Proceedings of CSCW '06. 2006*. ACM.

Taylor, A. G. (2000). *Wynar's Introduction to Cataloging and Classification*. Libraries Unlimited.

VanderWal, T. (2007). *Folksonomy Coinage and Definition*. Retrieved from http://vanderwal.net/folksonomy.html

KEY TERMS AND DEFINITIONS

Explicit Knowledge/Information: Knowledge that is external in the form of documents, graphs, tables etc.

Folksonomy: A set of terms generated by the collective action of an online community where users tag information or objects using their own terms or the terms recommended by others.

Generalization-Specialization Hierarchy: A set of concepts organized by specialization and generalization relationships into an inverted tree-like structure, such that concepts higher in the tree are broader and encompass the concepts lower in the tree

Knowledge Categories: Categories used to characterize the topics or areas of knowledge dealt with by documents.

Ontology: The set of the things that can be dealt with in a particular domain, together with their relationships.

Semantic Web: A vision of how the world-wide web could be more intelligent, based on metatagging the content together with the ability to inference automatically about different web objects are related to one another.

Tacit Knowledge/Information: Knowledge that is in people's heads and not externalized in documents or any other form.

Taxonomy: Any system of categories used to organize something, including documents, often less comprehensive than a thesaurus.

Thesaurus: Generally, a set of keywords used for indexing and information retrieval; in the Boeing case, various relationships, including synonymy, generalization, specialization, and related-to, were also included.

Training Sample: A set of documents or other pieces of text together with categories that are assigned to them to be used for training an automatic text classifier; there may be exactly one category assigned to each or there may be any number of categories, including zero, assigned to each piece of text.

This work was previously published in Encyclopedia of Knowledge Management, Second Edition, edited by David Schwartz, pp. 539-548, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 77

A Comparative Analysis of Activity-Based Costing and Traditional Costing Systems: The Case of Egyptian Metal Industries Company

Khaled Samaha
American University in Cairo, Egypt

Sara Abdallah
British University in Egypt, Egypt

EXECUTIVE SUMMARY

Today, organizational environments are increasingly characterized by an expanding use of advanced technologies. A company's management accounting system should capture the underlying technology, be consistent with corporate commitment to total quality and increased automation, and promote its efforts to compete on the basis of cost, quality, and lead time. However, the recent literature reveals that traditional cost accounting systems systematically introduce serious product cost distortions, which lead to inappropriate strategic decisions. Activity-Based Costing (ABC) represents an alternative paradigm that is giving more accurate and traceable cost information. The objective of this case is to illustrate the application of ABC method in a single manufacturing organization operating in the metal industry and to compare the results of ABC with volume based costing (traditional costing) method. The results of the application highlight the weak points of volume based costing which assigns factory overhead costs using direct labor-hours or machine-hours as a cost driver. As a result, volume-based costing under-costs low-volume product (i.e. products requiring fewer direct labor hours in total), while it over-costs high-volume products (i.e. products requiring more direct labor-hours in total), and thus, a product is subsidized at the expense of others. In cost accounting this is called cross-subsidization. However, activity-based costing traces overhead consumption by each product and thus provides a more accurate per-unit overhead cost.

DOI: 10.4018/978-1-4666-1945-6.ch077

1- ORGANIZATION BACKGROUND

Egyptian Metal Industries Co. (Metalco), an ISO 9002 certified company, was established in 1987 as the first Egyptian manufacturing company in the field of 'Hot Water Boiler and Radiators'. With headquarters in Egypt, Metalco provides superior quality products, serving both the domestic and international sectors. After over 20 years of long experience, and being always on top of all forms of new scientific improvement and technical developments, its research and development teams were able to further improve the production of hot water boilers to a more practical and higher efficient 'Hot Water Boilers', while maintaining the superior quality and designs that Metalco associates its name with. These new hot water boilers were given a new brand name "*Kinz*" which translates to the word "Treasure".

Heat exchangers are common components in many of the everyday devices. Central heating boilers and radiators all contain heat exchangers. Their purpose is to transfer heat from a hot liquid or gas to a colder one. In industry, steam is often used for heating and cold water for cooling. A variety of heat exchangers have been designed to suit the range of heating or cooling applications. Boiler and radiator heating systems are already green in many respects, since they are one of the cleanest heating systems around when it comes to fostering good indoor air quality - a big plus in green remodeling circles. Nevertheless, boiler or radiator heating system can still benefit from going green, especially when it comes to reducing the heating costs and improving the energy efficiency.

In 2005, the total world domestic boiler market was estimated at $ 10.9 billion and 10.46 million units, while it is expected to grow at a moderate rate over the next few years. The United Kingdom is the biggest market in both value and volume terms, followed by South Korea and Italy. Growth rates vary significantly between countries, with smaller markets generally offering higher growth potential. These are found in North Africa, the Middle East and Central Asia, with double-digit growth forecast in Kazakhstan, Tunisia, Morocco, Jordan and Turkey up to 2010. The growth expected in these markets occurs from a low base.

Today, Meltaco Company is one of Egypt's most mature manufacturing sectors operations, exclusively distributing and servicing more than 70,000 clients both in Egypt and in the world. Meltaco's goal is to exceed the expectations of every client by offering outstanding customer service, increased flexibility, and greater value, thus optimizing system functionality and improving operation efficiency. Meltaco associates are distinguished by their functional and technical expertise combined with their hands-on experience, thereby ensuring that our clients receive the most effective and professional service. Meltaco Company's policy is to provide a culture of continual improvement that is committed to exceed Customer expectations and requirements, by providing products and services of excellent quality. Customer support is number one at Meltaco Company

Meltaco Company is caring to develop its employee's staff and provide them with good work environment to achieve its goals and to maintain highly regarded performance level. Currently, it has a staff of 10 managers and 50 professionally trained maintenance personnel available 24 hours a day, 7 days a week to service all clients' needs. The company offers many different service plans. Meltaco's professional sales staff would assist the clients in choosing the best system of boilers or water heaters to meet their specific needs. The company warranties all its equipment, installation and spare parts. Customers are welcome to stop by the Spare Parts Department. While there, the company's professionally trained personnel will consult with each client to find the spare parts and equipment they need for their operation.

Meltaco produces a wide range of hot water boilers, including residential boilers, commercial boilers, swimming pool boilers, water heaters and laundry equipments. The most significant highly demanded product lines are Commercial Boiler (COM) and Residential Boiler (RES). Between those two products' lines, capacities range from 30,000 Kcal/hr up to 2,240,000 Kcal/hr. Each boiler type is produced in three types (1) Gas Boiler and (2) Oil Boiler and (3) Electric Boiler.

Seeking adequate operational information flow, the firm invests approximately 5% of their $30 million annual turnover in information systems adaptation. However, due to the information technology rapid advances, the company's management realizes the emerging need to update its current working systems. Top management support is imperative within this organization as they are the ones making the decisions. The organization seems to be under developed with regard to accounting information system (AIS) strategy planning; however they don't believe that it is a major problem. It is apparent that the president of the firm believes that investing in AIS will give a significant financial return. This shows a commitment by top management to forecast the long term benefits of AIS.

2- SETTING THE STAGE

Nour Khaled, the Factory Accountant, was sitting in a meeting with Glenn Hoddle, the General Manager of Egyptian Metal Industries Co. They were reviewing the March 2010 quarter's management accounts and discussing cost information on the company's two main product lines.

The general manager of the company was in full flight, frustrated as ever with the output from the company's costing system. 'How can it be? I just don't understand it. We've been losing business on the Residential Boilers line over the past quarter. Yet, your data indicates that it's our most profitable product. It doesn't seem logical to me.'

Glenn continued..........

Hisham Ibrahim (the Production Manager) insists that the Residential Boilers is so much harder to make than the commercial boilers range. How can it cost less to manufacture? Magda Khaled (the Sales Manager) is absolutely confident that, if they could reduce the commercial boilers selling price, they would see an even higher increase in sales. Are you sure that these costs are correct? How have you computed the overhead content of each product cost?'

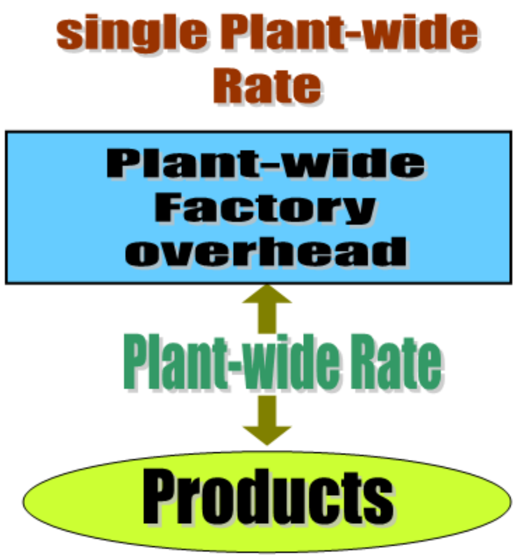
Khaled had covered this ground on many previous occasions with Hoddle and, yet again, set out to explain the operation of the company's costing system. In particular, she detailed the different steps in the allocation of overhead to the two product lines.

The overhead allocation process involves two separate steps. General Factory Overhead is spread across the two production departments (Assembly and Test) on the basis of the direct labor cost incurred in each department. The relevant portions of General Factory Overhead are then added to the separate overhead costs of the Assembly and Test departments before being divided by the total amount of direct labor hours required in each department. This exercise produces an overhead rate per direct labor hour for each production department. We then determine factory overhead rate for the entire production (i.e. plant-wide rate). Finally, we go on to apply that overhead per direct labor hour to the individual product's usage of direct labor hours (see Figure 1). We have used the same approach for years - long before we introduced the current products. No one has ever queried it. It's always worked before,..'

Before Khaled could continue, Hoddle interrupted her. It was clear that Hoddle was far more than impressed with this detailed and complicated explanation.

'Hmmm... Nour, plain common sense tells me that it's time to look again at the way we tackle costing. I was talking to Romani George at Liverpool Products limited last week and he said that his company had brought in something called XYZ, or was it ABC - some kind of new approach to costing. Would that be any simpler than that we do now? More important, do you think that it

Figure 1. Current overhead allocation in Metalco: A single plant wide rate



would produce any better quality cost data? Can you find out about it and let me know? We need fast answers on this - there are some good market opportunities out there, but we need to be in a position to respond to them. Come back to me as soon as you have had an opportunity to look into this.' In other words, the General Manager asked Khaled to provide a detailed analysis on *Overhead Allocation under the present costing system, and detailed computations on the effect of the implementation of an ABC approach on the allocation of total factory overhead.*

3- CHALLENGES FACING THE ORGANIZATION

A potential problem with this company's costing system is that overhead is calculated based upon direct labor cost, which is not the driving force behind the incurrence of the indirect costs in this company. Further, decision making regarding the company's product lines is not supported adequately to be of use to management.

Because volume-based costing uses an aggregate rate rather than a specific one, it loses a lot of information and control over product costs. A selected volume-based cost driver (e.g. direct labor-hours) might have little to do with how factory overhead costs accumulate. This can cause not only product price distortions, but it can also decrease the incentives for department managers to control product costs.

In traditional costing systems, direct materials and direct labor are the only cost components that can be traced directly to the product. Traditional costing systems were designed in the late nineteenth and early twentieth centuries, where the business environment was far less automated and the product line heterogeneity was less common. In this early era, the direct materials and direct labor truly reflect the primary components of the production costs, indirect overhead costs were relatively low, and the distortions arising from inappropriate overhead allocations were not significant. Information processing costs were high, and it was therefore difficult to justify more sophisticated overhead allocation (Xiong et al., 2008). However, today with far more automated production environment and companies produce a wide range of products, direct labor represents only a small fraction of total costs, and overhead indirect costs are far more of considerable importance. Such increase in indirect costs causes the overhead component to be the truly production prime costs. Simplistic overhead allocations using a declining direct labor base cannot be justified, particularly when information processing costs are no longer a barrier to introducing more sophisticated systems.

The increase in overhead cost portions distorts the product costing as traditional costing systems aggregate these overhead costs to product lines, and apply the aggregated overhead on a basis not necessarily consistent with the resources consumed by the product lines. Traditional overhead allocations are based on volume related bases (based on direct labor, direct materials, or machine hours), which may not be the driving force of the overhead activities. However, many organizational resources exist for activities that are unrelated to physical volume. Non-volume-related activities include probable support activities such as material handling, material procurement, set-ups, and production scheduling and first-item inspection activities. Also, the consumption of resources in production may differ extensively among product lines. Some costs or departmental cost pools may apply to a single or a few product lines, rather than to all products. Traditional product cost systems, which assume that products consume all resources in proportion to their production volumes, thus report distorted product costs (Hicks, 2006). Cooper and Kaplan (1988) claimed that the commonly used systematic overhead allocation methods distorts production costs, resulting in under-costing of low volume products and over-costing of high volume products, and thus, producing unreliable product costs to be utilized by management in attempt to control costs in the current highly competitive markets.

According to Cooper (1989), the traditional product costing misrepresentation could be regard to three prominent causes: (1) the failure to use direct costing where it is economically "feasible". In spite of the fact that advances in information technology have greatly reduced the costs associated with tracking and assigning costs to specific products and services, many companies under-utilize this technology. Rather than using meters to measure different products' energy consumptions, motion sensors to detect the amount of time employees spend at given tasks, and bar codes to help calculate the costs associated with moving specific materials and finished products, the majority of companies allow these costs to lapse into a general, catch-all category called 'overhead'; (2) the failure to segregate overhead costs into unique cost pools. All costs that cannot be classified as direct materials or direct labor the company categorizes into one cost pool, or at best a mere handful of cost pools. No attempt is made to separate overhead costs into common overhead cost categories. As a result, the company's ability to associate overhead costs with specific products, services, customers, or some other equally important attribute is blunted.(3) The failure to apply appropriate cost allocation bases. Just as there is often only one or a very limited number of overhead cost pools, so too there is often only one or a very limited number of cost drivers used to allocate the overhead costs. Frequently, direct labor hours or direct labor dollars are used to allocate the overhead costs. Critics have long charged that such a practice is antiquated and outmoded as labor costs as a percentage of product, and even service, are rapidly decreasing. For many of today's manufacturers direct labor costs represent less than 10 per cent of the product's total cost. Therefore, say these critics, the allocation of over-head based on direct labor is likely to result in distorted product costs.

In light of the above, Metalco Company may be operating with distorted product cost information, and thus likely to make incorrect pricing and product mix decisions. In particular, the company's general manager fears are stemming from the fact that the company may believe it is incapable of matching its competitors' prices on the high volume business, and thus likely to become increasingly committed to chasing the illusory profits that its cost accounting system suggests are available on the low volume business. Invariably the company ends up with low market share, high per unit overhead costs, and large losses. Ironically, therefore, it is often the company itself, or more precisely the accounting function, that strikes the mortal blow.

4- CASE DESCRIPTION

Nour left the General Manager's office, suppressing an inward groan. She had heard of ABC - activity based costing - but she was uncertain whether it would bring any benefits to the company's costing system.

Her main concern was that any change to the company's costing system would be painful to introduce and involve her in a considerable amount of extra work.

For the next month, Nour threw herself into acquiring as much knowledge as possible on ABC. She obtained copies of a couple of new text-books and scanned the ABC case study examples to see if they were relevant to the company. She contacted the accountant at Liverpool to learn from their experience in ABC implementation.

Nour noticed that an activity-based costing (ABC) system is a complex cost system that provides detailed and accurate information about each product's cost and thus each product's profitability. The ABC system is very expensive and time-consuming system to implement due to the high cost of collecting, measuring, and recording production information, especially for her company that is producing many different products that require hundreds of various activities. For a large company like Metalco, it might take millions of dollars and many years to develop and implement an ABC system. Despite its high cost, ABC systems have many benefits to offer.

The Advantages of an Activity-Based Costing System

An activity-based costing system has the following advantages:

Accurate product cost and product and customer profitability measurement: An ABC system provides accurate and detailed product information that can be utilized by management in decision making process. It can help to make better-informed management decisions concerning product pricing, product volumes, market segments, product lines, and target costing.

- Better-informed strategic decisions: The information about cost drivers for each activity enables management to make better-informed decisions concerning product design, customer support, etc.
- Production improvement: An ABC system provides information helpful in improving production processes.
- Better product cost information: An ABC system provides a lot of information necessary for strategic budgeting, planning, and product pricing decisions. It also takes into consideration cost of unused capacity and measures it at all levels (i.e. product, batch, and facility-levels).

Mechanics of ABC implementation

Nour recognized that there are five main steps involved in the implementation of ABC. These steps consist of:

1 identifying activities;
2 determining the costs of activities;
3 identifying cost drivers for each activity;
4 determining the annual capacity for the activity;
5 calculating the activity overhead rate.

The first step in implementing ABC is the identification of activities. Activities are best viewed as groupings of related actions. For example, the activity called 'machine set-up' is the outcome of a variety of individual actions. These actions might include cleaning or flushing the machine of any residue remaining from the previous batch that was run, adjusting the feed and speed of the machine for the new batch's requirements, and inspecting the first item produced to ensure that it conforms to the product's standards for fitness.

The second step in implementing ABC involves determining the costs of activities. Continuing with the machine set-up example used above, all the costs involved in executing machine set-ups need to be accumulated.

The third step in the implementation of ABC involves the selection of cost drivers. Actually, this third step consists of two sub stages.

The first-sub stage requires the identification of what are commonly called first-stage drivers. First-stage drivers trace the costs of inputs or resources into the cost pools that comprise each activity; remember an activity consists of a number of company-defined interrelated actions. It is possible, for instance, that a company may wish to include as a set-up cost the costs of moving material from inventory storage to the shop floor. Consequently, it is for this reason that an activity may consist of a number of cost pools. The appropriateness of these first-stage drivers has an important influence on the accuracy of the calculated product costs. Poorly chosen, first-stage drivers will likely jeopardize the usefulness and accuracy of the ABC system. The cost distortions that may arise will ripple through the ABC system and potentially be exaggerated by the more prone to error second-stage drivers. It is to this reason that companies may wish to trace the consumption of resources directly (as opposed to allocating them) to cost pools.

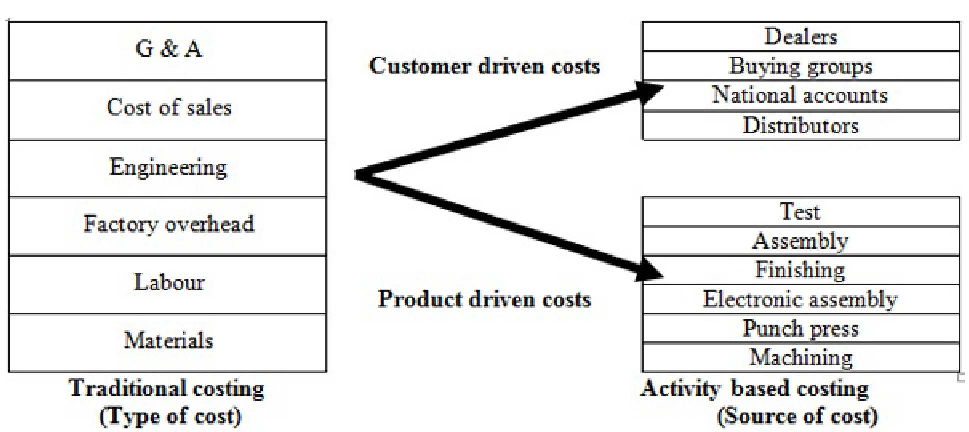
Second-sub stage drivers are used to allocate the cost pool overheads to products. The cost driver chosen should relate to the product's consumption of the particular activity's cost pool.

The fourth stage in the implementation of ABC requires the determination of the annual capacity for each activity. For example, if set-up hours are used as the cost driver for assigning the costs related to machine set-ups, then the company must determine the number of hours it has made available and therefore committed to this activity.

The fifth stage of ABC implementation involves the calculation of activity over-head rates. Let us say for the sake of illustration that 15,000 hours of machine set-up have been made available for the year at a cost of $ 600,000. Then the overhead rate for the set-up activity is $40 per set-up hour. This $40 rate will be used not only to cost products, but can also be used for the purpose of budgeting as well.

Activity-based costing focuses on the relative differences in assigning support costs. It uses the experience of operations people, and quantitative information, where available, to develop causal relationships between work drivers (what causes work to be done) and work to be done (activities). It allows the value of each activity to be assessed, identifies the factors that cause the activity to happen, and gives a true representation of the unit of measure of those activities (Figure 2).

Figure 2. ABC looks at what drives the cost



5- COMPARATIVE ANALYSIS AND PRACTICAL INSIGHTS: TRADITIONAL VERSUS ABC

To better understand the differences between the traditional and activity-based costing and how those differences affect product profitability analysis as well as management decisions, we will look at the Egyptian Metal industries company (Metalco) which has the data shown in Table 1 for the two products.

Using a Volume-Based Costing System

Under traditional costing, the factory overhead cost is assigned based on direct labor-hours (DLH). The factory overhead rate is determined as follows:

Factory Overhead Rate per DLH =

$$\frac{Total\ Factory\ Overhead}{Total\ DLH}$$

In this case, the factory overhead rate per DLH = $600,000 / 50,000 = $12 per DLH (i.e. plant-wide rate). Using this information, we can determine the total amount of factory overhead assigned to each product and the factory overhead cost per unit for both products (see Table 2).

Now we can calculate the unit margin profit for each product and determine how profitable the products are (see Table 3).

Table 1. Cost data for the two major products

	RES	COM
Production Volume	10,000	5,000
Selling Price	$78.00	$130.00
Unit prime cost	$21.00	$35.00
Direct labor-hours	30,000	20,000
Direct labor-hours per unit	3	4
Budgeted factory overhead	**$600,000**	
Budgeted direct labor-hours	**50,000**	

Table 2. Factory overhead cost per unit under traditional costing

	RES	COM
Factory overhead cost per DLH	**$12.00 per DLH**	
Direct labor hours	30,000	20,000
Factory overhead assigned	$360,000	$240,000
Production volume (in units)	10,000	5,000
Factory overhead cost per unit	$36.00	$48.00

Note: $360,000 = $12.00 x 30,000 and $240,000 = $12.00 x 20,000.

Table 3. Unit margin profit for each product under traditional costing

		RES	COM
Unit Selling Price	A	$78.00	$130.00
Less: Unit Prime Cost	B	($21.00)	($35.00)
Less: Unit Overhead Cost	C	($36.00)	($48.00)
Unit Margin	**D=A-B-C**	**$21.00**	**$47.00**
Unit Margin Percentage	**E=D/A × 100%**	**26.92%**	**36.15%**

Note C: $36.00 = $12.00 x 3 hrs and $48.00 = $12.00 x 4 hrs.

Therefore, using the volume-based costing we can see that the product COM is much more profitable than the product RES. However, let's use the activity-based costing and see if we can get the same or similar results.

Using an Activity-Based Costing (ABC)

Nour was able to gather some thoughts on other possible approaches to the apportionment of overheads. At last, she felt that he was ready for another meeting with the General Manager. Over the two weeks, Nour has assembled information which she believes will be relevant to the implementation of an ABC approach. In particular, she has reviewed the content of total factory overhead (including Assembly and Test departments' overheads) and considered how to set up appropriate cost pools and identify cost drivers.

Nour has identified the activities, costs, and activity consumption cost drivers shown in Table 4.

Nour also collected the activity data for each product (see Table 5).

Using the total cost for each activity and the total amount of activity cost driver we can determine the activity cost rate (see Table 6).

As we can see from Table 6, one set-up costs $294.12, one machine-hour costs $7.41, one inspection-hour costs $10, and one packing order costs $5.36. Now, let's calculate the per-unit cost of each product manufactured by Metalco Company (see Table 7).

Table 4. Activities, costs, and activity consumption cost drivers

Activity	Budgeted Cost	Cost Driver
Machine set-ups	$100,000	Number of set-ups
Machine running	$400,000	Machine-hours
Inspection	$70,000	Inspection-hours
Packing	$30,000	Number of packing-orders
Total	**$600,000**	

Table 5. Activity data for each product

Cost Driver	RES	COM	Total
Number of set-ups	80	260	340
Machine-hours	18,000	36,000	54,000
Inspection-hours	3,000	4,000	7,000
Number of packing-orders	2,100	3,500	5,600

Table 6. Activity cost rate

Cost Driver	Cost ($)	Activity Amount	Activity Rate ($)
	A	B	C=A/B
Number of Set-ups	100,000	340	294.12
Machine-hours	400,000	54,000	7.41
Inspection-hours	70,000	7,000	10.00
Number of packing-orders	30,000	5,600	5.36

The allocation of factory overhead using activity-based costing is summarized in Figure 3.

Using the obtained data we can perform the product profitability analysis; that is, we will determine the unit margin profit of each product (see Table 8).

According to the data presented in Table 8, the COM product is actually $22.57 ($37.18 – $14.61) less profitable than the RES product. The unit margin percentage for the COM product is lower than the one for the RES product when using the activity-based costing. Therefore, the management of Metalco Company might consider changing the production process of the COM product (e.g. increase batch-size to decrease the cost of machine set-ups per item produced) and increasing the selling price (i.e. as long as the increase in selling price would not substantially decrease product sales).

Using this example we can see that activity-based costing provides a lot of information about production activities and how they affect the cost of products, which the volume-based costing system does not explain.

Let's compare the product profitability analyses under volume-based and activity-based costing (see Table 9).

As we can see from Table 9, the product per-unit cost and per-unit margin profit differ under the volume-based and activity-based costing. The reason is that activity-based costing traces overhead consumption by each product and thus provides a more accurate per-unit overhead cost.

On the other hand, volume-based costing assigns factory overhead costs using direct labor-hours or machine-hours as a cost driver. As a result, volume-based costing under-costs the low-volume products (i.e. products requiring fewer direct labor hours in total), while it over-costs high-volume products (i.e. products requiring more direct labor-hours in total). In our example, the production of RES product and COM product required 30,000 and 20,000 direct labor-hours, respectively. Therefore, the overhead per unit cost

Table 7. Per-unit cost of each product manufactured under ABC

RES (10000 units)				
Activity Cost Driver	**Activity Rate**	**Activity Amount**	**Total Over-head**	**Overhead per Unit**
	A	B	C=A x B	D = C /10,000
Number of Set-ups	$ 294.12	80	23,529.60	$ 2.35
Machine-hours	7.41	18,000	133,380.00	13.34
Inspection-hours	10.00	3,000	30,000.00	3.00
Number of packing-orders	5.36	2,100	11,256.00	1.13
Total				*19.82*
COM (5000 units)				
Activity Cost Driver	**Activity Rate**	**Activity Amount**	**Total Over-head**	**Overhead per Unit**
	A	B	C=A x B	D = C /5,000
Number of Set-ups	$ 294.12	260	$ 76,471.20	$ 15.29
Machine-hours	7.41	36,000	266,760.00	53.35
Inspection-hours	10.00	4,000	40,000.00	8.00
Number of packing-orders	5.36	3,500	18,760.00	3.75
Total				**80.39**

Figure 3. Factory overhead using activity-based costing

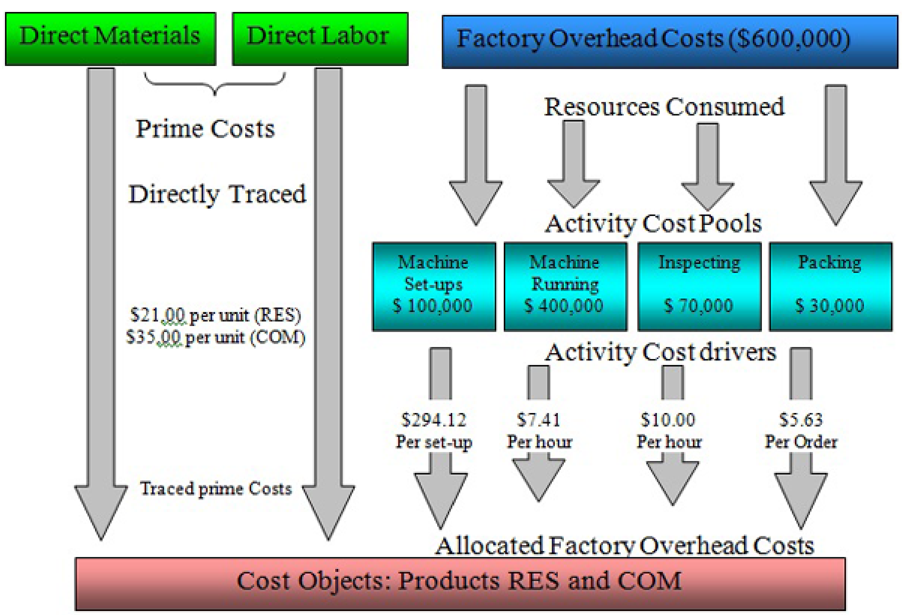


Table 8. Unit margin profit for each product under ABC

		RES	COM
Unit selling price	A	$78.00	$130.00
Less: Unit prime cost	B	($21.00)	($35.00)
Less: Unit overhead cost	C	($19.82)	($80.39)
Unit margin	**D = A-B-C**	**$37.18**	**$14.61**
Unit margin percentage	**E=D/A x 100%**	**47.67%**	**11.24%**

for products RES and COM were overstated and understated, accordingly.

Using the data obtained under the activity-based costing, for example, we can see that the overhead per-unit cost of product RES is overstated by $16.18 (i.e. $16.18 = $36.00 – $19.82) under the volume-based costing because the production of RES requires more direct labor-hours (in total). Thus, a product – COM in this case – is subsidized at the expense of others. In cost accounting this is called cross-subsidization.

Table 9. Product profitability analyses under volume-based and activity-based costing

	Volume-based costing		Activity-based costing	
	RES	**COM**	**RES**	**COM**
Unit selling price	$78.00	$130.00	$78.00	$130.00
Less: Unit prime cost	($21.00)	($35.00)	($21.00)	($35.00)
Less: Unit overhead cost	($36.00)	($48.00)	($19.83)	($80.39)
Unit margin	**$21.00**	**$47.00**	**$37.18**	**$14.61**
Unit margin percentage	**26.92%**	**36.15%**	**47.67%**	**11.24%**

REFERENCES

Cooper, R. (1989). You need a new cost system when…. *Harvard Business Review*, (Jan-Feb): 77–82.

Cooper, R., & Kaplan, R. S. (1988). How cost accounting systematically distorts product costs. *Management Accounting Quarterly*, 20-27. London.

Searcy, D. L. (2004). Using activity-based costing to assess channel/customer profitability. *Management Accounting Quarterly,* Winter, 51-60.

Thomas, M. F., & Mackey, J. (2006). Supply chain management: Monitoring strategic partnering contracts with activity-based measures. *Management Accounting Quarterly,* Fall, 1-10.

Thorne, H., & Gurd, B. (1999). Activity-based costing: Improved product costing or activity management? *Advances in Management Accounting*, *8*, 173–194.

Turney, P. B. B. (1992). What an activity-based cost model looks like. *Journal of Cost Management,* Winter, 54-60.

Xiong, Y., Su, W., & Lin, T. W. (2008). Activity-based costing popularity in China. *Cost Management,* May/June, 40-48.

KEY TERMS AND DEFINITIONS

Activity Cost Driver Rate: Is calculated by dividing activity expenses by the total quantity of the activity cost driver (e.g. machine set up expenses divided by total number of machine set up hours).

Activity-Based Costing (ABC): Is a product costing system when a company allocates factory overhead costs to activity centers (e.g. machine set ups, running machines) and then uses activity cost drivers to allocate factory overhead costs to individual products or services

Actual Costing: Is a product costing system when a company measures actual costs of direct materials, direct labor, and factory overhead. Actual costing system is rarely used because it does not provide accurate information on a timely basis: many costs can be measured only at the end of the production, and some actual costs fluctuate a lot leading to potential errors in price recording.

An activity Cost Driver: Is a quantity of activities needed to produce a product (e.g. 'machine set ups' activity center will have set-up hours as the activity cost driver).

Job Costing: Is a product costing system when costs are accumulated by specific job orders (e.g. Job Order XX2, Job Order 02357) and assigned to batches of products. In other words, manufacturing costs are assigned to specific jobs: specific customers, specific orders, specific projects, specific contracts, etc. Job costing is often used by small and medium-sized firms. Also, it is often used in the following industries: professional services (e.g. medical, legal), advertising agencies, construction, shipbuilding, custom equipment/furniture manufacturing, etc.

Normal Costing: Is a product costing system when a company measures the actual costs of direct materials and direct labor, but uses predetermined factory overhead rates to measure the factory overhead cost for a period. In other words, throughout the production time, the company measures and records the actual costs of direct materials and direct labor used, but it estimates a portion of factory overhead to be assigned to the product(s) (i.e. factory overhead applied). Normal costing system provides a timely cost estimate of a product or batch of products.

Process Costing: Is a product costing system when costs are accumulated by departments or processes (e.g. Printing Department, Assembling Department) and assigned to a large number of homogenous, identical products. In other words, manufacturing costs are assigned to each process in each manufacturing department: assembling costs (including direct materials, direct labor, and factory overhead) in the assembling department; cost of printing (direct labor, direct materials, and factory overhead) in the printing department, etc. Process costing is usually used by companies characterized by continuous mass production (i.e. firms that produce one or a few homogenous products). Process costing is often used in the following industries: textiles, food processing, automobile manufacturing, and electronics.

Standard Costing: Is a product costing system when a company measures all costs – direct materials, direct labor, and factory overhead – using standard quantities and costs.

Volume-Based Costing: Also called traditional costing, is a product costing system when a company allocates factory overhead costs to a single cost pool (e.g. factory overhead) and then uses volume based cost drivers to allocate factory overhead costs to individual products or services. The company uses volume-based cost drivers that depend on number of units manufactured. Cost bases (or drivers) often used are: labor hours, machine hours, labor costs, etc.

An easy way to remember the relationship between products: activity cost drivers, and resources, is to recall that products consume activities and activities consume resources.

This work was previously published in Cases on Business and Management in the MENA Region: New Trends and Opportunities, edited by El-Khazindar Business Research and Case Center, pp. 41-53, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 78
Complex Real-Life Supply Chain Planning Problems

Behnam Fahimnia
University of South Australia, Australia

Mohammad Hassan Ebrahimi
InfoTech International Company, Iran

Reza Molaei
Iran Broadcasting Services, Iran

ABSTRACT

Supply chain planning concerns the selection of strategies and methodologies to facilitate the optimal flow of material from raw material suppliers to end-users through procurement, production and distribution activities. Supply chain (SC) implementation has significant impacts on the financial performance of manufacturing and distribution companies. Developing real-life SC models with centralised planning naturally leads to complex models which are difficult to solve optimally. This chapter firstly presents a comprehensive review on the current literature of SC planning and optimisation and classifies the published models based on their complexity. Next, a mixed-integer non-linear formulation is presented for modelling complex real-life SC planning problems which accommodates the identified gaps in the current literature. Evaluation of the available tools and techniques for the optimisation of the proposed SC model will conclude this chapter.

1. INTRODUCTION TO SUPPLY CHAIN PLANNING

A supply chain (SC) can be defined as an integrated system synchronising a series of interrelated business processes in order to: (1) acquire raw materials and parts, (2) transform these to finished products (adding value), and (3) distribute final products to retailers or final customers. A SC plan facilitates the flow of information among SC participants including suppliers, manufacturers, distributors, retailers and final customers (Min & Zhou, 2002). Today, under the threat of increasing competition, firms must use effective planning and optimisation models and algorithms, decision support systems (DSS), and computerised analysis tools to improve

DOI: 10.4018/978-1-4666-1945-6.ch078

their operational performance along SCs (Yilmaz & Çatay, 2006). For this, researchers and practitioners have developed and implemented many SC planning models and methods for optimising their entire operations to win business (M. Chen & Wang, 1997).

The two core partial optimisation problems in the SC network are production planning and distribution planning problems. In production planning, decisions regarding hiring and firing of labour, regular time and overtime production, subcontracting, as well as machine capacity levels are made for a definite planning horizon (i.e. usually a one year period). Distribution planning decisions, on the other hand, pertain to determining which facility(ies) would cater to the demands of which market(s) (Mohamed, 1999). In a conventional SC, independent manufacturers, wholesalers, and retailers are separate business entities seeking to maximise their own profits although this goal is known to eventually produce profit for the system as a whole (Barbarosoglu & Ozgur, 1999). However, it is now widely acknowledged that production and distribution decisions are mutually related problems and need to be dealt with simultaneously in an integrated manner (Park, Choi, & Kang, 2007).

Unfortunately, developing real-life integrated SC models with centralised planning naturally leads to complex, large-scale models which are difficult to solve optimally. For this reason many alternative solution techniques developed in the literature are only able to provide near optimal solutions for small and medium-size integrated models (Barbarosoglu & Ozgur, 1999). This chapter presents a comprehensive review on the current literature of SC planning and optimisation. To accommodate the identified gaps in the current literature a mathematical formulation is presented for modelling complex real-life SC planning problems.

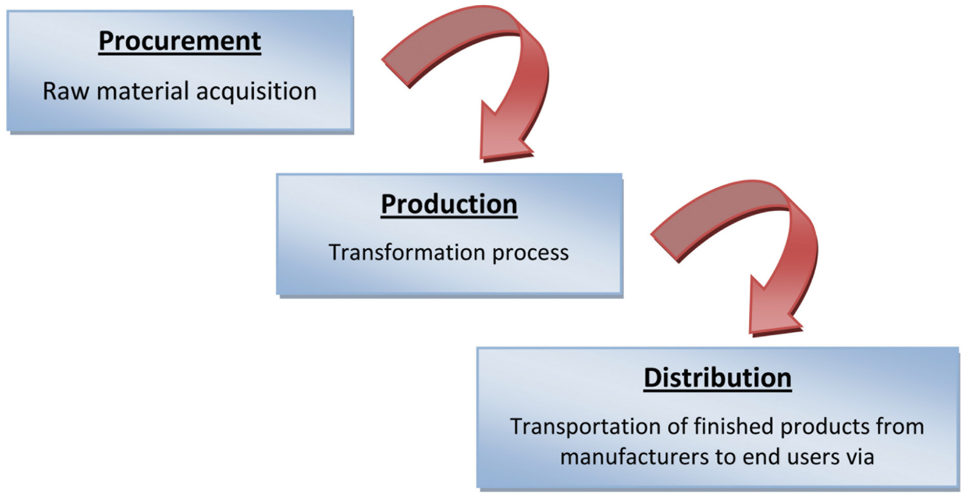
2. SUPPLY CHAIN PLANNING PROBLEM

A SC network includes procurement, manufacturing and distribution organisations working together to profitably provide the right product to the right customer at the right time (see Figure 1). Supply planning concerns with the selection of strategies and methodologies by the participating organisations (i.e. procurement, manufacturing and distribution) for satisfying the SC's short and long-term objectives (Gavirneni, 2006). From this definition, a SC manager is responsible for numerous decision makings such as the amount of raw material to purchase, production planning and inventory control issues, as well as transportation planning and warehousing decisions. In this chapter, SC is referred to as a production-distribution network in which the procurement activities are not incorporated and the SC network includes a set of manufacturers, warehouses and end-users.

SC planning is generally accomplished in 3 stages (Torabi & Hassini, 2009). The first stage is the strategic level or long-term planning where the SC configuration is determined (e.g. the location of manufacturing plants and warehouses). The objective at the second stage—tactical level or mid-term planning—is to determine the procurement, production and distribution quantities in order to minimise the overall SC costs while satisfying the customer demands. Finally, the third stage is the operational level or short-term planning where the day-to-day activities are managed according to the work plan drawn at the tactical level.

Over the last two decades, the importance of SC planning and optimisation at tactical and operational levels has been recognised as a competitive advantage for the growing production/distribution firms. There is also a growing recognition that these companies may not be able to compete without a global integration of all activities across SC (Sarmiento & Nagi, 1999). It is widely acknowledged that in order to achieve a globally

Figure 1. Configuration of SC network



optimised SC, manufacturing enterprises need to integrate and synchronise the activities in procurement, production and distribution networks for the coordination among the SC participants (Iannone, Miranda, & Riemma, 2007).

3. SIGNIFICANCE OF SUPPLY CHAIN PLANNING

Research indicates that SC planning and optimisation has raised significant interests for both researchers and practitioners over the past few years (Fahimnia, Luong, & Marian, 2008). There can be a number of major forces behind this trend, but there are two main issues which can perfectly justify the call for the implementation of effective and efficient integrated SC optimisation models (Fahimnia, Luong, & Marian, 2010). Firstly, the opportunity of an integrated management of a complex real-life SC positively affects the profitability of all the SC members (i.e. suppliers, manufacturers, and distributers). Secondly, effective SC planning reduces the lead-times and offers the quick response to market changes and therefore reduces the propagation of unexpected and undesirable events through the network (Guillén, Mele, Bagajewicz, Espuña, & Puigjaner, 2005; Persson & Olhager, 2002). These are the requirements of today's competition in order to stay in the market. This section discusses the two aforementioned factors.

3.1 Supply Chain Profitability

SC cost reduction has undoubtedly been one key issue why SCM has continually attracted much attention over the last few years. Optimising the supply chain performance is vital for the companies that are constantly examining areas where they can cut costs and improve profit margins while maintaining customer satisfaction (Koo, Adhitya, Srinivasan, & Karimi, 2008). Recently, manufacturing, distribution, and service industries have realised the cost savings that can be achieved by better planning and management of complex SC systems (Muriel & Simchi-Levi, 2003). Table 1 outlines some of the past success stories in the implementation of supply chain cost-saving plans (Aether-Consulting, 2009; Alloquor-Consulting, 2009; Bruss, 2005; Moduslink, 2009; The-cost-reduction-consultant, 2008).

Table 1 shows that the optimisation of complex SCs has become very attractive for both manu-

Table 1. SC success stories

Organisation	Overall cost savings achieved through the SC optimisation
Mazda Motor of America	$1.3M saving in annual SC costs through improving the logistics operation for the transportation of Mazda automobiles to the US dealers.
hp company	Savings of $24M per annum through the improvement of the SC transport management system
Nissan (North America Headquarters)	Annual cost reduction of 5%-10% through process and operations improvement
Pioneer Electronics	Annual savings of $1.5M (45%) through conducting a reverse logistics process improvement
Henry Schein	$1.3M savings in 6 months through the optimisation of the distribution network
Zara fashion store chain	28% rise in SC profits (equals to $621M) in nine month
Indiana Hospital	$3M cost reduction through cutting supply expenses
JLG Industries, Inc. (Producer of mobile lifts and excavators)	Potential savings opportunity of $11.4M identified (15.2% of total sourceable spend)
Advocate Health Care	$30M reduction (about 10%) in food, drug and supply costs in 18 months
Custom Building Products	12% reduction in annual transportation costs through reducing the number of carriers by 75%
Warner/Elektra/Atlantic (an AOL Time Warner company)	15% reduction in annual SC costs through logistics cost and service optimisation
Sur La Table	Reducing annual operating expenses by $1.2M through redesigning the distribution operation

facturers and service providers due to the opportunities it offers for improving the organisations' financial performance. It has been shown in the past that such financial enhancements are achieved when the decision-makings are integrated and coordinated among likeminded entities participating in the SC network (Pitty, Li, Adhitya, Srinivasan, & Karimi, 2008). In general, effective integration of procurement, production, and distribution activities across a SC network results in the following cost savings:

- **Reduction of production costs:** Production costs are the sum of fixed costs of opening and operating different manufacturing plants as well as the variable costs associated with production of various products at different plants. Implementation of integrated SC plans can contribute to the reduction of variable costs which may include regular-time production, overtime production, outsourcing as well as storage costs at stack buffers.
- **Reduction of distribution costs:** Distribution costs are the sum of fixed costs of opening and operating the warehouses and the variable costs of transporting finished goods from manufacturing plants to the end users through the warehouses. Implementation of integrated SC plans can contribute to the reduction of variable costs including transportation costs from plants to warehouses and from warehouses to end-users as well as penalty costs incurred for not being able to satisfy the customer demands at end-users.
- **Reduction of inventory costs:** In a complex SC, products and their components are manufactured in a number of different facilities and therefore inventory costs make up a significant proportion of total network costs (Fahimnia, Luong, &

Marian, 2009; Kaminsky & Kaya, 2008). Reduction of inventory costs through effective SCM contributes to a significant decrease in overall SC costs.

3.2 Quicker Response to Market Changes

The second issue on the significance of SC planning and optimisation is the quicker response to market changes and thus better customer satisfaction. To remain competitive by quickly responding to market changes, many companies started breaking traditional organisational barriers which has led to the high-level cooperation among different functional departments and introduced the concept of integrated SCM in 90s (Muriel & Simchi-Levi, 2003). Supply chain management which has attracted much interest thus far requires the effective use of information and communication systems to be widely implemented for providing data access to all SC participants (Muriel & Simchi-Levi, 2003). The result of this cooperation is appeared in form of lead-time reduction, an unavoidable element of today's time-based competitions. Sajadieh, Jokar and Modarres (2009) cites that an important benefit of the coordination is the more efficient and effective management of inventories across the entire SC which consequently leads to a shorter lead-time. In fact, the integrated management of a complex SC improves the SC information flow which naturally leads to an improved product flow (i.e. lower work-in-progress—WIP—inventories) and thereby shorter lead-time.

Two good examples of the application of SC strategies resulting in lead-time reduction are Seagate Technology and Intel Corporation (Alloquor-Consulting, 2009). Through the optimisation of SC plan, Seagate reduced its logistics cycle time by more than 50% and Intel reduced its WIP by over 35% while the average manufacturing cycle time was cut down from 10 to 4 days. In another case, a provider of communications products in United States could manage to create a 40% reduction in time to market (from 5 to 3 days), despite the production volume increases by more than 350% with a cumulative cost savings of $100M over 9 years (Moduslink, 2009).

The potential benefits of such lead-time reductions in SCM have been widely investigated and may include the following issues (Eskigun et al., 2005; Leng & Parlar, 2009; Ryu & Lee, 2003): (1) better responsiveness to market changes, (2) more accurate forecasts, (3) significant reduction of bullwhip effect throughout SC, (4) smaller order sizes, (5) reduction in WIP inventory and inventory of finished goods and (6) improved customer satisfaction.

4. REVIEW OF THE LITERATURE OF SUPPLY CHAIN PLANNING AND OPTIMISATION

The literature indicates that SC planning and optimisation has been an active research area over the last decade and many solutions have been made to solve the associated problems. The challenge in most of the SC planning problems at the tactical and operational level is to determine the operation policies to use so as to minimise the total cost or to maximise the profit of the SC while the SC configuration is assumed to be known. Mathematical techniques (e.g. linear programming, integer and mixed integer formulation) have been widely used to solve the small-scale SC planning problems and heuristics techniques, simulation and genetic algorithms have been recently applied to achieve near-optimal solutions with shorter computation time for larger-scale problems.

Many previous research works have attended the optimisation of a production plan in the supply chain context (Byrne & Bakir, 1999; J.-H. Chen & Ho, 2005; Ganesh & Punniyamoorthy, 2005; B. Kim & Kim, 2001; Pradenas, Penailillo, & Ferland, 2004; Tavakkoli-Moghaddam, Rabbani, Gharehgozli, & Zaerpour, 2007; Techawiboonwong & Yenradee, 2002; D. Wang & Fang, 1997; R.-C. Wang & Fang, 2001; R.-C. Wang & Liang, 2004). Many others have investigated the problems in the area of distribution planning (Amiri, 2006; Chopra, 2003; Jansen, Weert, Beulens, & Huirne, 2001; Liang, 2006; Lockett & Westwood, 1985; Raj, Nichols, Sterling, & Moynihan, 1992). However, in recent years simultaneous consideration of the characteristics and requirements of different functions in production and distribution networks to perform a global SC optimisation has attracted the attention of researchers and many models have been proposed in this direction. Sarmiento and Nagi (1999) cite that the basic idea behind all the proposed models in the area of SC planning is to simultaneously optimise decision variables of different functions that have traditionally been optimised sequentially (i.e. the optimised output of one stage becomes the input to the other).

Many reviews of the literature on the proposed SC models have been previously presented (Adrian E. Coronado Mondragon, Lyons, Michaelides, & Kehoe, 2006; Beamon, 1998; Donselaar, Kokke, & Allessie, 1998; Erenguc, Simpson, & Vakharia, 1999; Mary J. Meixell & Gargeya, 2005; Peidro, Mula, Poler, & Lario, 2009; Sarimveis, Patrinos, Tarantilis, & Kiranoudis, 2007, 2008; Sarmiento & Nagi, 1999; Thomas & Griffin, 1996; Van Donselaar & Sharman, 1997; Vidal & Goetschalckx, 1997). Despite the variety of the published reviews, our survey on the current literature indicates that there is no specific review on comparing the real world capabilities of the proposed SC models based on their level of complexity (multiplicity) and the solution approaches applied.

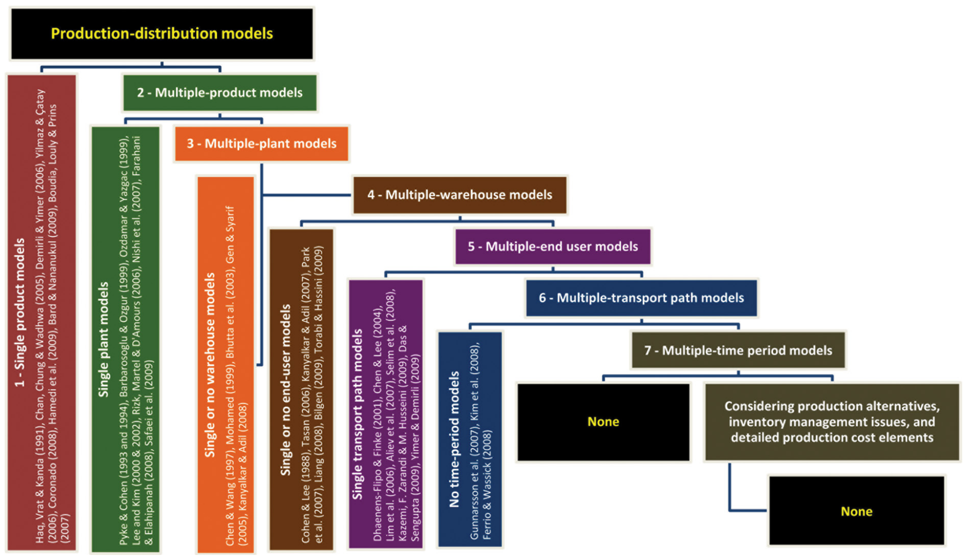
To evaluate the complexity of past research in the area, this section uses the following dimensions for classifying the past studies in the literature:

- Type of optimisation (i.e. total cost minimisation, profit maximisation, or multiple objective functions)
- Multiplicity of products to be produced at each manufacturing plant
- Multiplicity of geographically dispersed manufacturing plants
- Multiplicity of machine centres for the production of multiple products at each manufacturing plant
- A stack buffer at each manufacturing plant temporarily storing products prior to their shipment to the warehouses
- Multiplicity of transport paths from plants to end-users (direct/indirect shipment)
- Multiplicity of time-periods
- Multiplicity of production alternatives for each manufacturing plant
- Detailed production cost elements at aggregate level for each plant
- Inventory management issues considering the costs associated with WIP inventory and inventory of finished products at stack buffers and warehouses
- Shortage/penalty costs of not meeting demand forecast (also known as backlogging or backordering costs)
- Modelling and optimisation approaches

Based on the above-mentioned dimensions, all the published works in the context of SC planning and optimisation can be classified into the following 7 categories (refer to Figure 2).

1. Single-product models (Bard & Nananukul, 2009; Boudia, Louly, & Prins, 2007; Felix T. S. Chan, Chung, & Wadhwa, 2005; Coronado, 2008; Demirli & Yimer, 2006; Hamedi, Zanjirani Farahani, Husseini, & Esmaeilian, 2009; Haq, Vrat, & Kanda, 1991; Yilmaz & Çatay, 2006)
2. Multiple-product, single-plant models (Barbarosoglu & Ozgur, 1999; Elahipanah & Farahani, 2008; Lee & Kim, 2000, 2002; Nishi, Konishi, & Ago, 2007; Ozdamar & Yazgac, 1999; Pyke & Cohen, 1993, 1994; Rizk, Martel, & D'Amours, 2006; Safaei, Moattar Husseini, Z.-Farahani, Jolai, & Ghodsypour, 2009)

Figure 2. The classification of the literature of SC planning and optimisation



3. Multiple-product, multiple-plant, single or no warehouse models (Bhutta, Huq, Frazier, & Mohamed, 2003; M. Chen & Wang, 1997; Gen & Syarif, 2005; Kanyalkar & Adil, 2008; Mohamed, 1999)
4. Multiple-product, multiple-plant, multiple-warehouse, single or no end-user models (Bilgen, 2009; Cohen & Lee, 1988; Kanyalkar & Adil, 2007; Liang, 2008; Park, et al., 2007; Tasan, 2006; Torabi & Hassini, 2009)
5. Multiple-product, multiple-plant, multiple-warehouse, multiple-end user, single-transport path models (Aliev, Fazlollahi, Guirimov, & Aliev, 2007; C.-L. Chen & Lee, 2004; Dhaenens-Flipo & Finke, 2001; Lim, Jeong, Kim, & Park, 2006; Selim, Araz, & Ozkarahan, 2008)
6. Multiple-product, multiple-plant, multiple-warehouse, multiple-end user, multiple-transport path, no-time period models (Ferrio & Wassick, 2008; Gunnarsson, Rönnqvist, & Carlsson, 2007; Y. Kim, Yun, Park, Park, & Fan, 2008)
7. Multiple-product, multiple-plant, multiple-warehouse, multiple-end user, multiple-transport path, multiple time-period models (never studied in the literature)

Figure 2 illustrates the classification of previous works in the area of SC planning. Each level of this figure contains the related published works for that certain level of complexity. This figure can help identifying where the gaps are in the current literature of SC planning (refer to category 7 of this figure). Next section will discuss the research gaps identified.

5. IDENTIFIED GAPS IN THE LITERATURE OF SUPPLY CHAIN PLANNING

Despite the fact that the area of SC planning and optimisation has attracted much attention over the last few decades, there are still issues which have remained unaddressed. We classify the identified research gaps into two categories.

The first identified issue is "oversimplification" in the proposed models which is mainly due to the complexities associated with the optimisation of a complex real-life SC plan. Previous studies on the SC planning and optimisation focus on some SC principles, while pay less attention to the rest of the activities in the network. However, by actively pursuing only a subset of principles manufacturing/distribution firms will not succeed in achieving their expected improvements in the integrated SC performance (Muckstadt, Murray, Rappold, & Collins, 2001). Paying less attention to considering a realistic set of variables and constraints can preclude the previous models from functioning effectively in the real-life manufacturing/distribution scenarios. Therefore, there is a need to further extend the scope of the proposed studies in the way to develop real-world models concerning with the optimisation of complex integrated SC plans.

An evidence of oversimplification in the previously developed models is the scarcity of their real world implementation for solving complex large-scale SC planning problems. To demonstrate the applicability of a SC optimisation model, it is essential to test the algorithm on real world SC case problems incorporating a realistic range of variables and constraints (Blackhurst, Wu, & O'Grady, 2005). This helps analysing the impacts of various decisions on different SC entities (Biswas & Narahari, 2004). Although some of the former studies have shown the successful application of their models in the optimisation of simplified SC planning problems, the approaches have not been implemented in real-life large-scale case problems. in fact, a real world SC problem can be much larger in size and more complex in nature than the proposed models in the literature.

The second identified gap relates to the lack of appropriate and effective solution approaches for handling large-scale SC planning problems. Various techniques have been used in the literature to solve small and medium size SC planning problems ranging from mathematical programming, heuristics, simulation and knowledge based system to the latest fuzzy programming and genetic algorithms (Lair, 2008). Finding the optimal solution in a complex real-life SC planning problem with the use of the presented tools and approaches in the literature is almost impossible or subject to heavy computing overheads. Some of other presented approaches which might be able to handle larger-scale problems are not able to prove the optimality of the solutions found or do not have the potential to find the optimal solution on their own. Moreover, there is a wide range of tools and techniques available for the optimisation of SC models and a typical user may not be aware of the various tools available for solving a particular decision problem (Biswas & Narahari, 2004). Hence, there is a need to further enhance the quality and precision of the solution approaches for the optimisation of complex large-scale real-life SC planning problems.

The following 2 sections aim to address the identified research gaps in the current literature by developing a mathematical formulation for modelling a complex real-life SC planning problem and evaluating the alternative tools and techniques available for the optimisation of the developed mixed-integer non-linear model.

6. A COMPLEX LARGE-SCALE SUPPLY CHAIN PLANNING MODEL

The complex real-life SC planning problem considered in this research consists of a number of production plants producing a variety of product types during multiple time periods. The finished products are then distributed to a set of end-users through multiple transport routes in the sense that the products can be transported directly from manufacturing plants to end-users or indirectly through several warehouses. All the demand is to be satisfied at the end-users by the end of the planning horizon. The proposed SC plan aims the efficient utilisation of production allocation as well as transport, warehousing and material flow.

6.1 Key Performance Indicators

The first issue in developing a mathematical model is to identify the appropriate key performance indicators of the system. A vast amount of the literature has been published suggesting performance indicators for SC modelling and optimisation purposes (Bullinger;, Kühner, & Hoof, 2002; Gunasekaran, Patel, & Tirtiroglu, 2001; Lapide, 2000; S¨urie & Wagner, 2005; Stadtler & Kilger, 2005). However, all the proposed alternatives can fit within four main categories of *performance measures* that are applicable in most SC settings: (1) overall SC costs, (2) inventory management, (3) delivery performance (time-based) and (4) SC responsiveness (S¨urie & Wagner, 2005).

The literature on the SC planning models indicates that the cost-based value characteristics have been the most popular performance measures among both researchers and practitioners (Lair, 2008). These may include the minimisation of total cost, setup cost, delivery cost and penalty cost or the maximisation of overall SC profit. Obviously, the significance of cost value as the first SC performance indicator is due to the large interests of operation managers to observe the direct financial impacts on the overall SC performance (F. T. S. Chan, 2003). Likewise, the proposed model in this research aims to minimise the overall system costs as the main SC performance measure because cost-based optimisation can clearly reflect the SC efficiency due to its direct financial implication on the SC performance.

6.2 Model Assumptions

A set of assumptions are considered in the proposed SC planning model:

- Variety of products to be produced is known.
- Number, location and capacity of plants and warehouses are known.
- Number and location of end-users (customer zones) are known.
- Demand is deterministic and the aggregate demand for all types of products in the concerned periods is assumed to be known for the next planning horizon. The aggregate demand at each end-user is the total demand for each product which may be ordered by several individuals or retailers at end-users.
- Capacity limitations for regular-time and over-time production (capacity of machine centres), capacity of raw material supply, limitations in storage capacity at stack buffers and warehouses, and distribution capacities are known.
- All customer demands for each product have to be satisfied, sooner or later, during the planning horizon. A penalty cost will be incurred if the demand for a certain product at one period is decided to be backordered. The backordered demands are to be satisfied in the next periods before the end of the planning horizon.
- End-users or customer zones are the locations where products are delivered to the final customers with no holding capacity to store the products.
- Transportation costs are proportional to the transportation distances.
- Orders for subcontracting items are made by each plant at the start of each period and subcontracted items are sent directly to the stack buffer of the source plant (ordering plant) by the end of ordering period.

6.3 Parameters and Decision Variables

Before formulating the objective function, all the input parameters and decision variables must be clearly described. Indices used for mathematical formulation of the proposed model include product index (i), plant index (m), machine centre index (g), stack buffer index (b), warehouse index (w), end-user index (e) and time-period index (t).

6.3.1 Parameters

Parameters represent the input data for a system. Therefore, a parameter can be referred to as a variable with a fixed given value which is used as an input to the model (Lair, 2008). The parameters used for the purpose of mathematical modelling in this chapter are listed in Table 2.

6.3.2 Decision Variables

Decision variables are the outputs of the system or the variables in which the values need to be determined by the systems. The decision variables for the presented model in this chapter are listed in Table 3.

The integer variables are presented in Table 4.

6.4 Objective Function Formulation

This section describes the development of a mixed-integer nun-linear formulation model for a complex real-life SC plan. The proposed model deals with the trade-off between costs of inventories, production, transports and storage by integrating the Aggregate Production Plan and Distribution Plan. The objective function minimises the sum of the following costs: (1) production costs in regular-time, (2) production costs in over-time, (3) outsourcing costs, (4) inventory holding costs in plants, buffers, and warehouses, (5) transportation costs and (6) backlogging costs. The cost function (i.e. the objective function) is presented in Equation 1 (see Table 5).

The objective function in Equation 1 consists of 12 cost components:

- *Component 1 and 5* are the fixed costs of opening and operating plants and warehouses respectively. These are independent of the rate and quantities of production/distribution at plants/warehouses and may include hiring costs of building and facilities, amortisations of machines and tools, salaries of the managers, annual insurance payments among others.
- *Component 2* expresses the production costs in regular-time derived from Equation 2 (see Table 6).
- *Component 3* describes the over-time production costs derived from Equation 3 (see Table 7).
- *Component 4* is outsourcing or sub-contracting costs as an option for failing to manufacture the products in-house.
- *Component 6, 7, and 8* represent the inventory holding costs in plants, buffers and warehouses respectively.
- *Component 9, 10, and 11* express the transportation costs for the distribution of items from plants to end-users. This can be done directly from plants to end-users (expressed in component 11) or indirectly from plants to warehouses and from warehouses to end-users (expressed in components 9 and 10 respectively).
- *Component 12* stands for the shortage (penalty) costs incurred if backlogging occurs at the end-users.

6.5 Model Constraints

This section highlights all the realistic constraints faced in a real world SC plan including (1) the capacity constraints of manufacturing plants, buffers and warehouses, (2) the distribution capacity constraints of buffers and warehouses, (3) the demand satisfaction and backlogging constrains, (4) the inventory level constraints at the start/end of the planning horizon as well as (5) the variables constrains.

Table 2. Parameters

D_{iet}	Forecasted demand for product *i* at end-user e in period *t*
O_{mt}	Fixed costs of opening and operating plant *m* in period *t*
O'_{wt}	Fixed costs of opening and operating warehouse *w* in period *t*
H_{imt}	Unit WIP inventory holding cost for product *i* at plant m in period *t*
H'_{ibt}	Unit holding cost for finished product *i* at stack buffer *b* in period *t*
H''_{iwt}	Unit holding cost for finished product *i* at warehouse *w* in period *t*
HC_{ibt}	Holding capacity (maximum units) at stack buffer *b* for product *i* in period *t*
HC'_{iwt}	Holding capacity (units) at warehouse *w* for product *i* in period *t*
$Xmax_{imt}$	Maximum allowed WIP inventory for the finished product *i* carried in plant *m* at the end of period *t*
T_{ibwt}	Unit transportation cost of product *i* from buffer *b* to warehouse *w* in period *t*
T'_{iwet}	Unit transportation cost of product *i* from warehouse *w* to end-user *e* in *t*
T''_{ibet}	Unit transportation cost of product *i* from buffer *b* to end-user *e* in period *t*
RP_{imt}	Unit regular-time production cost of product *i* at plant *m* in period *t*
OP_{imt}	Unit over-time production cost of product *i* at plant *m* in period *t*
OS_{imt}	Unit outsourcing cost of product *i* ordered by plant *m* in period *t*
P_{igmt}	Processing time to produce a unit of product *i* on machine *g* at plant *m* in *t*
Q_{igmt}	Average time spent to produce a WIP unit of product *i* on machine *g* at plant *m* in *t*
L_{igmt}	Labour/hour cost for regular-time production of product *i* on machine *g* at plant m in period *t*
L'_{igmt}	Labour/hour cost for over-time production of product *i* on machine *g* at plant *m* in *t*
RM_{imt}	Cost of raw material for producing a unit of product *i* at plant *m* in period *t*
α_{imt}	Variable overhead costs of regular-time production of product *i* at plant *m* in *t*
β_{imt}	Variable overhead costs of over-time production of product *i* at plant *m* in *t*
SC_{iet}	Unit backordering (shortage) cost for product *i* at end-user *e* in period *t*
$Smax_{iet}$	Maximum amount of shortage permitted (maximum backorder) for product *i* at end-user *e* in period *t*
λ_{igmt}	Capacity hours for regular-time production of product *i* on machine *g* at plant *m* in period *t*
λ'_{igmt}	Capacity hours for over-time production of product *i* on machine *g* at plant *m* in period *t*
Y_{imt}	Capacity units of raw material supply for product *i* at plant *m* in period *t*
E_{ibt}	The distribution capacity at stack buffer *b* for product *i* in period *t*
E'_{iwt}	The distribution capacity at warehouse *w* for product *i* in period *t*
μ_{im}	WIP inventory level of product *i* in plant *m* at the start of planning horizon
μ'_{im}	WIP inventory level of product *i* in plant *m* at the end of planning horizon
η_{ib}	Inventory level of product *i* in stack buffer *b* at the start of planning horizon
η'_{ib}	Inventory level of product *i* in stack buffer *b* at the end of planning horizon
Π_{iw}	Inventory level of product *i* in warehouse *w* at the start of planning horizon
Π'_{iw}	Inventory level of product *i* in warehouse *w* at the end of planning horizon
$\rho_{bm} = \begin{cases} 1 & If\ stack\ buffer\ b\ is\ located\ in\ plant\ m \\ 0 & Otherwise \end{cases}$	

Table 3. Decision variables

I_{imt}	Quantity of product *i* produced in regular-time at plant *m* in period *t*
I'_{imt}	Quantity of product *i* produced in over-time at plant *m* in period *t*
I''_{imt}	Quantity of product *i* outsourced by plant *m* in period *t*
J_{ibwt}	Quantity of product *i* shipped from buffer *b* to warehouse *w* during period *t*
J'_{iwet}	Quantity of product *i* shipped from warehouse *w* to end-user *e* during period *t*
J''_{iwet}	Quantity of product *i* shipped directly from buffer *b* to end-user *e* during *t*
X_{imt}	WIP inventory amount of product *i* at plant *m* at the end of period *t*
Y_{ibt}	Inventory amount of finished-product *i* at buffer *b* at the end of period *t*
Z_{iwt}	Inventory amount of product *i* at warehouse *w* at the end of period *t*
S_{iet}	Quantity of product *i* backordered at end-user *e* at the end of period *t*

Table 4. Integer variables

$F_{ibwt} = \begin{cases} 1 & \text{If product i is shipped from bufer b to warehouse w at period t} \\ 0 & \text{Otherwise} \end{cases}$	
$F'_{iwet} = \begin{cases} 1 & \text{If product i is shipped from warehouse w to end-user e at period t} \\ 0 & \text{Otherwise} \end{cases}$	
$F''_{ibet} = \begin{cases} 1 & \text{If product i is shipped from buffer b to end-user e at period t} \\ 0 & \text{Otherwise} \end{cases}$	
$G_{mt} = \begin{cases} 1 & \text{If plant m operates in period t} \\ 0 & \text{Otherwise} \end{cases}$	
$G'_{wt} = \begin{cases} 1 & \text{If warehouse w is open in period t} \\ 0 & \text{Otherwise} \end{cases}$	
$d_{iet} = \begin{cases} 1 & \text{If demand for product i at end-uder e is not met at period t} \\ 0 & \text{Otherwise} \end{cases}$	

Table 5. Equation 1

$$
\begin{aligned}
MinZ = & \sum_t \sum_m O_{mt} G_{mt} + \sum_i \sum_m \sum_t I_{imt} RP_{imt} G_{mt} + \sum_i \sum_m \sum_t I'_{imt} OP_{imt} G_{mt} \\
& + \sum_i \sum_m \sum_t I''_{imt} OS_{imt} G_{mt} + \sum_t \sum_w O'_{wt} G'_{wt} + \sum_i \sum_m \sum_t H_{imt} X_{imt} G_{mt} \\
& + \sum_i \sum_b \sum_m \sum_t H'_{ibt} Y_{ibt} \rho_{bm} G_{mt} + \sum_i \sum_w \sum_t H''_{iwt} Z_{iwt} G'_{wt} \\
& + \sum_i \sum_b \sum_m \sum_w \sum_t J_{ibwt} T_{ibwt} \rho_{bm} G_{mt} G'_{wt} F_{ibwt} + \sum_i \sum_w \sum_e \sum_t J'_{iwet} T'_{iwet} G'_{iwet} F'_{iwet} \\
& + \sum_i \sum_b \sum_m \sum_w \sum_t J''_{ibet} T''_{ibet} \rho_{bm} G_{mt} F''_{ibet} + \sum_i \sum_e \sum_t S_{iet} SC_{iet} d_{iet}.
\end{aligned}
$$

Table 6. Equation 2

$$\sum_{i}\sum_{m}\sum_{t} I_{imt}\ RP_{imt}\ G_{mt} = \sum_{i}\sum_{m}\sum_{t}\left[I_{imt}\ G_{mt}\left(\sum_{g} P_{igmt}\ \ L_{igmt} + RM_{imt} + \alpha_{imt}\right)\right]$$

Table 7. Equation 3

$$\sum_{i}\sum_{m}\sum_{t} I'_{imt}\ OP_{imt}\ G_{mt} = \sum_{i}\sum_{m}\sum_{t}\left[I'_{imt}\ G_{mt}\left(\sum_{g} P_{igmt}\ \ L'_{igmt} + RM_{imt} + \beta_{imt}\right)\right]$$

Capacity constraints of plants:

- Restrictions on raw material supply:

$$I_{imt} + I'_{imt} \leq \gamma_{imt}\ G_{mt}\ \ \forall i,m,t \tag{4}$$

- Production capacity constraint for regular-time and over-time production:

$$If \quad \sum_{i}\sum_{g} Q_{igmt} = \frac{\sum_{i}\sum_{g} P_{igmt}}{2} \quad \forall i,m,t \text{ Then:}$$

$$\begin{aligned} &\sum_{i}\sum_{g} P_{igmt}\ \ I_{imt} + \sum_{i}\sum_{g} P_{igmt}\ \ I'_{imt} + \\ &\sum_{i}\sum_{g} Q_{igmt}\left(X_{imt} - X_{im(t-1)}\right) \\ &\leq \sum_{i}\sum_{g} G_{mt}\ \left(\lambda_{igmt} + \lambda'_{igmt}\right) \quad \forall i,m,t \end{aligned} \tag{5}$$

- Limitation on WIP inventory amount to be carried at each plant:

$$X_{imt} \leq X_{imt}^{Max}\ \ \forall i,m,t \tag{6}$$

Capacity constraints of stack buffers:

- Stack buffer capacity restriction:

$$Y_{ibt} \leq HC_{ibt}\ \forall i,b,t \tag{7}$$

- Inventory balance at stack buffers (see Table 8).

Capacity constraints of warehouses:

- Warehouse capacity restriction:

$$Z_{iwt} \leq HC'_{iwt}\ \ \forall i,w,t \tag{9}$$

- Inventory balance at warehouses (see Table 9).

Distribution capacity limits at stack buffers:

$$\sum_{w} J_{ibwt} + \sum_{e} J''_{ibet} \leq\ E_{ibt}\ \rho_{bm}\ G_{mt}\ \ \forall i,b,m,t \tag{11}$$

Table 8. Equation 8

$$Y_{ibt} - Y_{ib(t-1)} = G_{mt}\left[(I_{imt} + I'_{imt} + I''_{imt}) - \rho_{bm}\ (\sum_{w} J_{ibwt}\ G'_{wt} + \sum_{e} J''_{ibet})\right]\ \forall i,b,m,t$$

Table 9. Equation 10

$$Z_{iwt} - Z_{iw(t-1)} = \sum_{b} J_{ibwt}\, \rho_{bm}\, G_{mt}\, G'_{wt}\, F_{ibwt} - \sum_{e} J'_{iwet}\, G'_{wt}\, F'_{iwet} \quad \forall i,w,t$$

Distribution capacity constraint at warehouses:

$$\sum_{e} J'_{iwet} \le G'_{wt} E'_{iwt} \;\; \forall i,w,t \tag{12}$$

Demand satisfaction and backlogging constraints at end-users:

- Demand satisfaction constraint: The total amount of production and outsourcing for every product at all plants must meet the forecasted demand for that product at the end of planning horizon.

$$\sum_{m}\sum_{t}\left(I_{imt} + I'_{imt} + I''_{imt}\right) = \sum_{e}\sum_{t} D_{iet} \;\; \forall i \tag{13}$$

- Maximum allowed shortage at end-users:

$$S_{iet} \le S^{Max}{}_{iet} \;\; \forall i,e,t \tag{14}$$

- Inventory balance at end-users (see Table 10)

Constraints on the start/end inventory levels:

- Constraint on the WIP inventory level (Equation 16) and the inventory level of finished products at stack buffers (Equation 17) and warehouses (Equation 18) at the start and end of the planning horizon (t=0 and t=T):

$$\sum_{t=0} X_{imt} = \mu_{im} \;\; \& \sum_{t=T} X_{imt} = \mu'_{im} \;\; \forall i,m \tag{16}$$

$$\sum_{t=0} Y_{ibt} = \eta_{ib} \;\; \& \sum_{t=T} Y_{ibt} = \eta'_{ib} \;\; \forall i,b \tag{17}$$

$$\sum_{t=0} Z_{iwt} = \varphi_{iw} \;\; \& \sum_{t=T} Z_{iwt} = \varphi'_{iw} \;\; \forall i,w \tag{18}$$

Non-negativity restriction for all decision variables:

$$I_{imt}, I'_{imt}, I''_{imt}, X_{imt} \ge 0 \;\; \forall i,m,t \tag{19}$$

$$J_{ibwt} \ge 0 \; \forall i,b,w,t \tag{20}$$

$$J'_{iwet} \ge 0 \;\; \forall i,w,e,t \tag{21}$$

$$J''_{ibet} \ge 0 \;\; \forall i,b,e,t \tag{22}$$

$$Y_{ibt} \ge 0 \; \forall i,b,t \tag{23}$$

$$Z_{iwt} \ge 0 \; \forall i,w,t \tag{24}$$

$$S_{iet} \ge 0 \; \forall i,e,t \tag{25}$$

Table 10. Equation 15

$$\sum_{w} J'_{iwet}\, G'_{wt} + \sum_{b} J''_{ibet}\, \rho_{bm}\, G_{mt} = D_{iet} - S_{iet}.\, d_{iet} + S_{ie(t-1)}\, .\, d_{ie(t-1)} \;\; \forall i,e,t$$

7. TOOLS AND TECHNIQUES FOR THE OPTIMISATION OF COMPLEX SUPPLY CHAIN MODELS

All optimisation processes concern with the minimisation or maximisation of a function in which the values of the decision variables are selected systematically from a set of alternatives. The origin of many optimisation techniques which we use today dates back to the second World War (e.g. Simplex method) when the massive logistical issues raised by huge armies having millions of men and machines (Chinneck, 2004a; Dantzig, 2002). Today, optimisation methods are used almost everywhere from business, industry, and government to engineering and computer sciences because demand for optimisation modelling arise as regularly as they did during World War II (Chinneck, 2004a). In general, enterprises are continuously searching for new optimisation tools and techniques in order to obtain better results in terms of cost reduction, capital turnover, and quality improvement (Castro-Lacouture, Medaglia, & Skibniewski, 2007).

Over the last few decades, a large number of sophisticated optimisation problems have raised in the area of supply chain management to improve the decisions of SC planners. These types of optimisation promise the improvements in company's SC performance through reduced SC system costs, lower inventories, shorter lead-times, increased manufacturing throughput, as well as better return on assets. Among these, optimisation of complex SC models is one of the most popular problems (Paksoy, Kürsat Gules, & Bayraktar, 2007). The survey on the current literature indicates that several techniques and methodologies have been proposed in the past for the optimisation of integrated SC plan. Each of these techniques has strengths and weaknesses and can be helpful in solving limited types of problems. We classify these techniques into four categories: mathematical techniques, heuristics techniques, simulation, and genetic algorithms.

The essence of all mathematical techniques (e.g. linear programming, non-linear programming, mixed integer programming, and Lagrangian Relaxation) is to use the mathematical language for describing the essential aspects/functions of an actual system (Chinneck, 2004a, 2004b; Jordan & Smith, 2002). A range of mathematical techniques have been adopted in the literature to solve SC planning problems including *Linear Programming models* (M. Chen & Wang, 1997; Kanyalkar & Adil, 2005, 2007, 2008; Liang, 2008), *Mixed Integer Programming models* (Bhutta, et al., 2003; Bilgen, 2009; C.-L. Chen & Lee, 2004; Das & Sengupta, 2009; Demirli & Yimer, 2006; Dhaenens-Flipo & Finke, 2001; Ferrio & Wassick, 2008; Gunnarsson, et al., 2007; Hamedi, et al., 2009; Haq, et al., 1991; Y. Kim, et al., 2008; Mohamed, 1999; Paksoy, et al., 2007; Rizk, et al., 2006; Selim, et al., 2008; Tsiakis & Papageorgiou, 2008; H. Yan, Yu, & Edwin Cheng, 2003), and *Lagrangian Relaxation models* (Barbarosoglu & Ozgur, 1999; Jayaraman & Pirkul, 2001; Nishi, et al., 2007; Syam, 2002).

Mathematical programming models have been demonstrated to be useful analytical tools in optimising decision-making problems such as those encountered in SC planning (Geoffrion, 1976; Xu, He, & Gen, 2009). Proposed in 1947, Linear programming (LP) is only applicable when all of the underlying models of the real world processes are linear (Chinneck, 2004a; Dantzig, 2002). LP has been widely used in the past for solving constrained optimisation problems. In mathematical modelling, some of the variables may be real/fractional values and some others may only take integer values (0, 1). Mixed Integer Programming (MIP) is used when a combination of these two is required to be included in a model. MIP may be used in two forms: Mixed Integer Linear Programming (MILP) and Mixed Integer Nonlinear Programming (MINLP)—which is much harder to solve (Chinneck, 2004b). In Lagrangian Relaxation methodology, those constraints that make the problem hard to solve are removed from the model and are put into the objective function. A weight is then assigned to each constraint representing a penalty value when a constraint is not fully satisfied. In this way, the objective function value represents the solution feasibility in the sense that not only it includes the cost factors, but it is also a function of the constrain satisfaction. All the mathematical techniques are fully matured and are guaranteed producing the optimal solution for a certain problem types (Lair, 2008).

However, the following issues may restrict the application of mathematical modelling in solving complex real-world SC planning problems:

1. Developing a mathematical algorithm is not always easy. The complexity of formulating a mathematical model increases exponentially with the increase in the number of decision variables and model constraints (Lair, 2008; Méndez, Cerdá, Grossmann, Harjunkoski, & Fahl, 2006). Most of the SC planning problems are from a complex nature with a high number of variables and constraints involved and hence mathematical optimisation methods like LP and MIP may not be very effective in solving real-world SC planning problems (Jordan & Smith, 2002; Lair, 2008). In many cases, a complex SC planning problem may not even have a discrete feasible solution space to be searched by a mathematical technique (Chinneck, 2004b). Hence, mathematical techniques may be only suitable for solving small to medium size SC planning problems with limited number of variables and constraints.
2. Due to the exponential growth of the model complexity, immense mathematical models are NP-Hard problems which are very hard to solve (Lair, 2008; Park, et al., 2007). According to George Dantzig (2002), the father of linear programming, high computer memory and long CPU time is required in order to process complex mathematical algorithms. This can make it almost impossible to deal with a real-life large-scale SC planning problem using mathematical techniques unless the complex problems are simplified to a large extent in order to avoid the above-mentioned issues. This could have been one of main reasons for the oversimplification in the SC models developed in the past.

These restrictions of mathematical techniques have brought up the use of experience-based techniques known as heuristics for finding feasible solutions in large-scale SC planning problems. The primary advantage of heuristics methods is that they are able to rapidly find a solution that is hoped to be close to the optimal. One of the most popular heuristics which have been widely used in the literature is Simulated Annealing (Chinneck, 2004c). There have been some attempts for the use of heuristics methods in solving SC problems using other heuristics techniques (Cohen & Lee, 1988; Coronado, 2008; Ozdamar & Yazgac, 1999; Pyke & Cohen, 1993, 1994; Yeh, 2005; Yilmaz & Çatay, 2006).

There are, however, two reasons why heuristics techniques are not always the preferred technique for the optimisation of complex SC plans. Firstly, heuristics techniques cannot guarantee the optimality of the solutions found. Chinneck cites in his "Practical Optimization" book that as oppose to mathematical techniques, heuristics do not promise the optimal solutions but usually (not always) are able to give a good acceptable solution in solving complex problems (Chinneck, 2004c). Secondly, heuristics techniques are not effective tools in locating the global optimal solution when dealing with a complex problem with a vast search space, which is always the case in solving SC planning problems (Lair, 2008; Xing, 2006). Simulated Annealing, for instance, has limited capability in exploring large search spaces due to examining only one point of the search space at a time (Xing, 2006).

Simulation modelling in the area of supply chain management has been also used to monitor the performance of a real system so that the associated problems are diagnosed, SC activities are evaluated and possible solutions to the problems are suggested (Tarokh & Golkar, 2006). Based on these capabilities, simulation techniques can be ideal for reproducing the behaviours of complex systems (Tarokh & Golkar, 2006). Many previous studies have shown the capability of simulation modelling in SC modelling and optimisation (Jain, Ervin, Lathrop, Workman, & Collins, 2001; Jain, Workman, Collins, & Ervin, 2001; Lee, Cho, Kim, & Kim, 2002; Lee & Kim, 2000, 2002; Lim, et al., 2006; Moyaux, Chaib-draa, & D'Amours, 2004; Ritchie-Dunham, Morrice, Scott, & Anderson, 2000; Safaei, et al., 2009; Sarker et al., 2005; Williams & Gunal, 2003; T. Yan, Luh, Weidong, & Narimatsu, 2003). Due to the stochastic nature of today's SCs, simulation can be a highly effective tool in visualising the variations in uncertain SC factors and making effective and efficient business decisions (Williams & Gunal, 2003).

While simulation modelling is capable of describing various real and detailed situations, the two downsides of simulation modelling can demonstrate the limited application of this methodology for the optimisation of complex SC planning models (Lair, 2008; Park, et al., 2007). Firstly, simulation modelling is not able to guarantee the optimality of the developed solution. In many cases, even generating the near optimal solution cannot be promised using this method. Secondly, simulation software packages are in one hand expensive to purchase and on the other hand very time-consuming to analyse the auto-generated outputs.

Introduced by John Holland (1975), Genetic Algorithm (GA) techniques are based on the evolutionary ideas of natural selection and genetics. GAs are basically designed to simulate processes in a natural system necessary for evolution, particularly those following the principles laid down by Charles Darwin about the survival of the fittest (Ganesh & Punniyamoorthy, 2005). GAs are general-purpose search methods which are able to achieve a good balance between exploration and exploitation of the search space (Fahimnia, 2006). GA has been proven to be a highly effective and efficient tool in solving complex optimisation problems and some of their successful applications have already been reported in the context of SC planning (Aliev, et al., 2007; Altiparmak, Gen, & Lin, 2005; Altiparmak, Gen, Lin, & Karaoglan, 2007; Altiparmak, Gen, Lin, & Paksoy, 2006; Felix T. S. Chan, et al., 2005; Elahipanah & Farahani, 2008; Gen & Syarif, 2005; Kazemi, Fazel Zarandi, & Moattar Husseini, 2009; Liang, 2008; Park, et al., 2007; Syarif, Yun, & Gen, 2002; Tasan, 2006; Yeh, 2006; Yimer & Demirli, 2009).

The main advantages of using GA techniques for solving large optimisation problems are their robustness, searching flexibility and their evolutionary nature (Xing, 2006). GAs are capable of searching large and complicated search spaces in which the optimal solution is achieved by the convergence of the fitness function as the number of evolutions raises (Fahimnia, 2006; Holland, 1975; Lair, 2008; Marian, 2003; Sobhi-Najafabadi, 2002). The GA nature-based guided searching mechanism is well-structured and coordinated that enables this technique to produce the optimal solution for complex problems faster than any other optimisation tool; while for problems that are not overly difficult, other methods may find the solution even faster than GAs (Fahimnia, 2006; Haupt & Haupt, 2004; Lair, 2008; Marian, 2003; Sobhi-Najafabadi, 2002). However, the optimisation procedure (i.e. the evaluation of fitness function) in GA is performed on a large population of solutions (generally between 100 to 500 individuals) and this may require parallel processors when dealing with larger-scale SC planning problems (Haupt & Haupt, 2004).

REFERENCES

Aether-Consulting. (2009). Supply chain consulting and product cost reduction project services. Retrieved 31 August, 2010, from http://www.aetherconsulting.com/cons_supply_chain.htm

Aliev, R. A., Fazlollahi, B., Guirimov, B. G., & Aliev, R. R. (2007). Fuzzy-genetic approach to aggregate production-distribution planning in supply chain management. *Information Sciences*, *177*(20), 4241–4255. doi:10.1016/j.ins.2007.04.012

Alloquor-Consulting. (2009). Supply chain consultants - Logistics consultants.

Altiparmak, F., Gen, M., & Lin, L. (2005). A genetic algorithm for supply chain network design. Paper presented at the 35th International Conference on Computers and Industrial Engineering.

Altiparmak, F., Gen, M., Lin, L., & Karaoglan, I. (2007). A steady-state genetic algorithm for multi-product supply chain network design. *Computers & Industrial Engineering*, *56*(2), 521–537. doi:10.1016/j.cie.2007.05.012

Altiparmak, F., Gen, M., Lin, L., & Paksoy, T. (2006). A genetic algorithm approach for multi-objective optimization of supply chain networks. *Computers & Industrial Engineering*, *51*(1), 196–215. doi:10.1016/j.cie.2006.07.011

Amiri, A. (2006). Designing a distribution network in a supply chain system: Formulation and efficient solution procedure. *European Journal of Operational Research*, *171*(2), 567–576. doi:10.1016/j.ejor.2004.09.018

Barbarosoglu, G., & Ozgur, D. (1999). Hierarchical design of an integrated production and 2-echelon distribution system. *European Journal of Operational Research*, *118*(3), 464–484. doi:10.1016/S0377-2217(98)00317-8

Bard, J. F., & Nananukul, N. (2009). The integrated production-inventory-distribution-routing problem. Journal of Scheduling, 12(3), 257-280. doi: http://dx.doi.org/10.1007/s10951-008-0081-9

Beamon, B. M. (1998). Supply chain design and analysis: Models and methods. *International Journal of Production Economics*, *55*(3), 281–294. doi:10.1016/S0925-5273(98)00079-6

Bhutta, K. S., Huq, F., Frazier, G., & Mohamed, Z. (2003). An integrated location, production, distribution and investment model for a multinational corporation. *International Journal of Production Economics*, *86*(3), 201–216. doi:10.1016/S0925-5273(03)00046-X

Bilgen, B. (2010). Application of fuzzy mathematical programming approach to the production allocation and distribution supply chain network problem. *Expert Systems with Applications: An International Journal*, *37*(6), 4488–4495. doi:10.1016/j.eswa.2009.12.062

Biswas, S., & Narahari, Y. (2004). Object oriented modeling and decision support for supply chains. *European Journal of Operational Research*, *153*(3), 704–726. doi:10.1016/S0377-2217(02)00806-8

Blackhurst, J., Wu, T., & O'Grady, P. (2005). PCDM: A decision support modeling methodology for supply chain, product and process design decisions. *Journal of Operations Management*, *23*(3-4), 325–343. doi:10.1016/j.jom.2004.05.009

Boudia, M., Louly, M. A. O., & Prins, C. (2007). A reactive GRASP and path relinking for a combined production-distribution problem. *Computers & Operations Research*, *34*(11), 3402–3419. doi:10.1016/j.cor.2006.02.005

Bruss, J. (2005). Mastering the supply chain: Challenged with saving $30 million in 18 months, Illinois' largest IDN combine integrated business applications with organizationwide best practices.

Bullinger, H.-J., Kühner, M., & Hoof, A. V. (2002). Analysing supply chain performance using a balanced measurement method. *International Journal of Production Research*, *40*(15), 3533–3543. doi:10.1080/00207540210161669

Byrne, M. D., & Bakir, M. A. (1999). Producion planning using a hybrid simulation - Analytical approach. *International Journal of Production Economics*, *59*(1-3), 305–311. doi:10.1016/S0925-5273(98)00104-2

Castro-Lacouture, D., Medaglia, A. L., & Skibniewski, M. (2007). Supply chain optimization tool for purchasing decisions in B2B construction marketplaces. *Automation in Construction*, *16*(5), 569–575. doi:10.1016/j.autcon.2006.08.005

Chan, F. T. S. (2003). Performance measurement in a supply chain. *International Journal of Advanced Manufacturing Technology*, *21*(7), 534–548. doi:10.1007/s001700300063

Chan, F. T. S., Chung, S. H., & Wadhwa, S. (2005). A hybrid genetic algorithm for production and distribution. *Omega*, *33*(4), 345–355. doi:10.1016/j.omega.2004.05.004

Chen, C.-L., & Lee, W.-C. (2004). Multi-objective optimization of multi-echelon supply chain networks with uncertain product demands and prices. *Computers & Chemical Engineering*, *28*(6-7), 1131–1144. doi:10.1016/j.compchemeng.2003.09.014

Chen, J.-H., & Ho, S.-Y. (2005). A novel approach to production planning of flexible manufacturing systems using an efficient multi-objective genetic algorithm. *International Journal of Machine Tools & Manufacture*, *45*(7-8), 949–957. doi:10.1016/j.ijmachtools.2004.10.010

Chen, M., & Wang, W. (1997). A linear programming model for integrated steel production and distribution planning. *International Journal of Operations & Production Management*, *17*(6), 592–610. doi:10.1108/01443579710167276

Chinneck, J. W. (2004a). Chapter 1: Introduction. In Practical optimization: A gentle introduction. Ottawa, Ontario, Canada.

Chinneck, J. W. (2004b). Chapter 13: Binary and mixed-integer programming. In Practical optimization: A gentle introduction. Ottawa, Ontario, Canada.

Chinneck, J. W. (2004c). Chapter 14: Heuristics for discrete search: Genetic algorithms and simulated annealing. In Practical optimization: A gentle introduction. Ottawa, Ontario, Canada.

Chopra, S. (2003). Designing the distribution network in a supply chain. *Transportation Research Part E, Logistics and Transportation Review*, *39*(2), 123–140. doi:10.1016/S1366-5545(02)00044-3

Cohen, M. A., & Lee, H. L. (1988). Strategic analysis of integrated production-distribution systems: Models and methods Operations Research Society of America, 36(2), 216-228.

Coronado, J. L. (2008). An optimization model for strategic supply chain design under stochastic capacity disruptions. Doctor of Engineering, Texas A&M University, Texas. Retrieved from http://handle.tamu.edu/1969.1/85844

Coronado Mondragon, A. E., Lyons, A. C., Michaelides, Z., & Kehoe, D. F. (2006). Automotive supply chain models and technologies: A review of some latest developments. *Journal of Enterprise Information Management*, *19*(5), 551–562. doi:10.1108/17410390610703675

Dantzig, G. B. (2002). Linear programming. *Operations Research*, *50*(1), 42–47. doi:10.1287/opre.50.1.42.17798

Das, K., & Sengupta, S. (2009). A hierarchical process industry production-distribution planning model. *International Journal of Production Economics*, *117*(2), 402–419. doi:10.1016/j.ijpe.2008.12.003

Demirli, K., & Yimer, A. D. (2006). Production-distribution planning with fuzzy costs. Paper presented at the Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American.

Dhaenens-Flipo, C., & Finke, G. (2001). An integrated model for an industrial production-distribution problem. *IIE Transactions*, *33*, 705–715. doi:10.1080/07408170108936867

Donselaar, K. V., Kokke, K., & Allessie, M. (1998). Performance measurement in the transportation and distribution sector. *International Journal of Physical Distribution & Logistics Management*, *28*(6), 434–450. doi:10.1108/09600039810245085

Elahipanah, M., & Farahani, R. Z. (2008). A genetic algorithm to optimize the total cost and service level for just-in-time distribution in a supply chain. *International Journal of Production Economics*, *111*(2), 229–243. doi:10.1016/j.ijpe.2006.11.028

Erenguc, S. S., Simpson, N. C., & Vakharia, A. J. (1999). Integrated production/distribution planning in supply chains: An invited review. *European Journal of Operational Research*, *115*(2), 219–236. doi:10.1016/S0377-2217(98)90299-5

Eskigun, E., Uzsoy, R., Preckel, P. V., Beaujon, G., Krishnan, S., & Tew, J. D. (2005). Outbound supply chain network design with mode selection, lead times and capacitated vehicle distribution centers. *European Journal of Operational Research*, *165*(1), 182–206. doi:10.1016/j.ejor.2003.11.029

Fahimnia, B. (2006). Optimisation of manufacturing lead-time, using genetic algorithm. Thesis for the degree of Master of Engineering, University of South Australia, Adelaide.

Fahimnia, B., Luong, L., & Marian, R. (2008). Optimization/simulation modeling of the integrated production-distribution plan: An innovative survey. *WSEAS Transactions on Business and Economics*, *5*(3), 44–57.

Fahimnia, B., Luong, L., & Marian, R. (2009). Optimization and performance evaluation of an integrated production-distribution plan in supply chains. Paper presented at the ASME International Mechanical Engineering Congress & Exposition (IMECE2009) Florida, USA.

Fahimnia, B., Luong, L., & Marian, R. (2011). (in press). Genetic algorithm optimization of an integrated aggregate production-distribution plan. *International Journal of Production Research*.

Ferrio, J., & Wassick, J. (2008). Chemical supply chain network optimization. *Computers & Chemical Engineering*, *32*(11), 2481–2504. doi:10.1016/j.compchemeng.2007.09.002

Ganesh, K., & Punniyamoorthy, M. (2005). Optimization of continuous-time production planning using hybrid genetic algorithms-simulated annealing. [London, UK: Springer-Verlag Limited.]. *International Journal of Advanced Manufacturing Technology*, *26*, 148–154. doi:10.1007/s00170-003-1976-4

Gavirneni, S. (2006). Information centric optimisation of inventories in capacitated supply chains: 3 illustrative examples. In J. Geunes & P. M. Pardalos (Eds.), Supply chain optimization (illustrated ed., pp. 1-50). New York, NY: Springer Science + Business Media Inc.

Gen, M., & Syarif, A. (2005). Hybrid genetic algorithm for multi-time period production/distribution planning. *Computers & Industrial Engineering*, *48*(4), 799–809. doi:10.1016/j.cie.2004.12.012

Geoffrion, A. M. (1976). A guided tour of recent practical advances in integer linear programming. *Omega*, *4*(1), 49–57. doi:10.1016/0305-0483(76)90038-4

Guillén, G., Mele, F. D., Bagajewicz, M. J., Espuña, A., & Puigjaner, L. (2005). Multiobjective supply chain design under uncertainty. *Chemical Engineering Science*, *60*(6), 1535–1553. doi:10.1016/j.ces.2004.10.023

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management*, *21*(1/2), 71–87. doi:10.1108/01443570110358468

Gunnarsson, H., Rönnqvist, M., & Carlsson, D. (2007). Integrated production and distribution planning for Södra Cell AB. *Journal of Mathematical Modelling and Algorithms*, *6*(1), 25–45. doi:10.1007/s10852-006-9048-z

Hamedi, M., Zanjirani Farahani, R., Husseini, M. M., & Esmaeilian, G. R. (2009). A distribution planning model for natural gas supply chain: A case study. *Energy Policy*, *37*(3), 799–812. doi:10.1016/j.enpol.2008.10.030

Haq, A. N., Vrat, P., & Kanda, A. (1991). An integrated production-inventory-distribution model for manufacture of urea: A case. *International Journal of Production Economics*, *25*(1-3), 39–49. doi:10.1016/0925-5273(91)90129-H

Haupt, R., & Haupt, S. (2004). Practical genetic algorithms. In Discrete applied mathematics (Second edition ed., vol. 146, pp. 1-261). New Jersey: John Wiley & Sons, Inc. Publication.

Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor, MI: University of Michigan Press.

Iannone, R., Miranda, S., & Riemma, S. (2007). Supply chain distributed simulation: An efficient architecture for multi-model synchronization. *Simulation Modelling Practice and Theory*, *15*(3), 221–236. doi:10.1016/j.simpat.2006.10.004

Jain, S., Ervin, E. C., Lathrop, A. P., Workman, R. W., & Collins, L. M. (2001). Analyzing the supply chain for a large logistics operation using simulation. Paper presented at the Simulation Conference, 2001. Proceedings of the Winter.

Jain, S., Workman, R. W., Collins, L. M., & Ervin, E. C. (2001). Development of a high-level supply chain simulation model. Proceedings of the Winter Simulation Conference, 2001.

Jansen, D. R., Weert, A. v., Beulens, A. J. M., & Huirne, R. B. M. (2001). Simulation model of multi-compartment distribution in the catering supply chain. *European Journal of Operational Research*, *133*(1), 210–224. doi:10.1016/S0377-2217(00)00204-6

Jayaraman, V., & Pirkul, H. (2001). Planning and coordination of production and distribution facilities for multiple commodities. *European Journal of Operational Research*, *133*(2), 394–408. doi:10.1016/S0377-2217(00)00033-3

Jordan, D. W., & Smith, P. (2002). *Mathematical techniques: An introduction for the engineering, physical, and mathematical sciences* (3rd ed.). Oxford, UK: OUP.

Kaminsky, P., & Kaya, O. (2008). Inventory positioning, scheduling and lead-time quotation in supply chains. *International Journal of Production Economics*, *114*(1), 276–293. doi:10.1016/j.ijpe.2008.02.006

Kanyalkar, A., & Adil, G. (2005). An integrated aggregate and detailed planning in a multi-site production environment using linear programming. *International Journal of Production Research*, *43*(20), 4431–4454. doi:10.1080/00207540500142332

Kanyalkar, A., & Adil, G. (2007). Aggregate and detailed production planning integrating procurement and distribution plans in a multi-site environment. *International Journal of Production Research*, *45*(22), 5329–5353. doi:10.1080/00207540600632174

Kanyalkar, A., & Adil, G. (2010). A robust optimisation model for aggregate and detailed planning of a multi-site procurement-production-distribution system. *International Journal of Production Research*, *48*(3), 635–656. doi:10.1080/00207540802471272

Kazemi, A., Fazel Zarandi, M., & Moattar Husseini, S. (2009). A multi-agent system to solve the production–distribution planning problem for a supply chain: A genetic algorithm approach. *International Journal of Advanced Manufacturing Technology*, *44*(1), 180–193. doi:10.1007/s00170-008-1826-5

Kim, B., & Kim, S. (2001). Extended model for a hybrid production planning approach. *International Journal of Production Economics*, *73*(2), 165–173. doi:10.1016/S0925-5273(00)00172-9

Kim, Y., Yun, C., Park, S. B., Park, S., & Fan, L. T. (2008). An integrated model of supply network and production planning for multiple fuel products of multi-site refineries. *Computers & Chemical Engineering*, *32*(11), 2529–2535. doi:10.1016/j.compchemeng.2007.07.013

Koo, L. Y., Adhitya, A., Srinivasan, R., & Karimi, I. A. (2008). Decision support for integrated refinery supply chains: Part 2. Design and operation. *Computers & Chemical Engineering*, *32*(11), 2787–2800. doi:10.1016/j.compchemeng.2007.11.007

Lair, A. M. N. (2008). An integrated model for optimising manufacturing and distribution network scheduling. PhD Thesis, University of South Australia, Adelaide.

Lapide, L. (2000). What about measuring supply chain performance? AMR Research White Paper.

Lee, Y. H., Cho, M. K., Kim, S. J., & Kim, Y. B. (2002). Supply chain simulation with discrete-continuous combined modeling. *Computers & Industrial Engineering*, *43*(1-2), 375–392. doi:10.1016/S0360-8352(02)00080-3

Lee, Y. H., & Kim, S. H. (2000). Optimal production-distribution planning in supply chain management using a hybrid simulation-analytic approach. Paper presented at the Simulation Conference, 2000, Winter.

Lee, Y. H., & Kim, S. H. (2002). Production-distribution planning in supply chain considering capacity constraints. *Computers & Industrial Engineering*, *43*(1-2), 169–190. doi:10.1016/S0360-8352(02)00063-3

Leng, M., & Parlar, M. (2009). Lead-time reduction in a two-level supply chain: Non-cooperative equilibria vs. coordination with a profit-sharing contract. *International Journal of Production Economics*, *118*(2), 521–544. doi:10.1016/j.ijpe.2009.01.002

Liang, T.-F. (2006). Distribution planning decisions using interactive fuzzy multi-objective linear programming. *Fuzzy Sets and Systems*, *157*(10), 1303–1316. doi:10.1016/j.fss.2006.01.014

Liang, T.-F. (2008). Fuzzy multi-objective production/distribution planning decisions with multi-product and multi-time period in a supply chain. *Computers & Industrial Engineering*, *55*(3), 676–694. doi:10.1016/j.cie.2008.02.008

Lim, S. J., Jeong, S. J., Kim, K. S., & Park, M. W. (2006). A simulation approach for production-distribution planning with consideration given to replenishment policies. *International Journal of Advanced Manufacturing Technology*, *27*(5), 593–603. doi:10.1007/s00170-004-2208-2

Lockett, A. G., & Westwood, J. B. (1985). Distribution planning in a turbulent environment. *European Journal of Operational Research*, *19*(1), 33–40. doi:10.1016/0377-2217(85)90305-4

Marian, R. (2003). Optimisation of assembly sequences using genetic algorithms. PhD thesis, University of South Australia, Adelaide.

Mary, J. Meixell, & Gargeya, V. B. (2005). Global supply chain design: A literature review and critique. *Transportation Research Part E, Logistics and Transportation Review*, *41*(6), 531–550. doi:10.1016/j.tre.2005.06.003

Méndez, C. A., Cerdá, J., Grossmann, I. E., Harjunkoski, I., & Fahl, M. (2006). State-of-the-art review of optimization methods for short-term scheduling of batch processes. *Computers & Chemical Engineering*, *30*(6-7), 913–946. doi:10.1016/j.compchemeng.2006.02.008

Min, H., & Zhou, G. (2002). Supply chain modeling: past, present and future. *Computers & Industrial Engineering*, *43*(1-2), 231–249. doi:10.1016/S0360-8352(02)00066-9

Moduslink. (2009). Case study: Provider of communication products.

Mohamed, Z. M. (1999). An integrated production-distribution model for a multi-national company operating under varying exchange rates. *International Journal of Production Economics*, *58*(1), 81–92. doi:10.1016/S0925-5273(98)00080-2

Moyaux, T., Chaib-draa, B., & D'Amours, S. (2004). Multi-agent simulation of collaborative strategies in a supply chain. Paper presented at the Third International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2004.

Muckstadt, J. A., Murray, D. H., Rappold, J. A., & Collins, D. E. (2001). Guidelines for collaborative supply chain system design and operation. *Information Systems Frontiers*, *3*(7), 427–453. doi:10.1023/A:1012824820895

Muriel, A., & Simchi-Levi, D. (2003). Supply chain design and planning - Applications of optimization techniques for strategic and tactical models. In Lenstra, J. K., & Nemhauser, G. L. (Eds.), *Handbooks in operations research and management science* (*Vol. 11*, pp. 15–93). Elsevier. doi:10.1016/S0927-0507(03)11002-X

Nishi, T., Konishi, M., & Ago, M. (2007). A distributed decision making system for integrated optimization of production scheduling and distribution for aluminum production line. *Computers & Chemical Engineering*, *31*(10), 1205–1221. doi:10.1016/j.compchemeng.2006.10.006

Ozdamar, L., & Yazgac, T. (1999). A hierarchical planning approach for a production-distribution system. *International Journal of Production Research*, *37*(16), 3759. doi:10.1080/002075499190031

Paksoy, T., Kürsat Gules, H., & Bayraktar, D. (2007). Design and optimization of a strategic production-distribution model for supply chain management: Case study of a plastic profile manufacturer in Turkey. *Selçuk Journal of Applied Mathematics*, *8*(2), 83–99.

Park, B., Choi, H., & Kang, M. (2007). Integration of production and distribution planning using a genetic algorithm in supply chain management. In P. Melin, O. Castillo, E. G. Ramírez, & W. Pedrycz (Eds.), Analysis and design of intelligent systems using soft computing techniques (pp. 416-426).

Peidro, D., Mula, J., Poler, R., & Lario, F.-C. (2009). Quantitative models for supply chain planning under uncertainty: A review. *International Journal of Advanced Manufacturing Technology*, *43*(3), 400–420. doi:10.1007/s00170-008-1715-y

Persson, F., & Olhager, J. (2002). Performance simulation of supply chain designs. *International Journal of Production Economics*, *77*(3), 231–245. doi:10.1016/S0925-5273(00)00088-8

Pitty, S. S., Li, W., Adhitya, A., Srinivasan, R., & Karimi, I. A. (2008). Decision support for integrated refinery supply chains: Part 1. Dynamic simulation. *Computers & Chemical Engineering*, *32*(11), 2767–2786. doi:10.1016/j.compchemeng.2007.11.006

Pradenas, L., Penailillo, F., & Ferland, J. (2004). Aggregate production planning problem: A new algorithm. *Electronic Notes in Discrete Mathematics*, *18*, 193–199. doi:10.1016/j.endm.2004.06.031

Pyke, D. F., & Cohen, M. A. (1993). Performance characteristics of stochastic integrated production-distribution systems. *European Journal of Operational Research*, *68*(1), 23–48. doi:10.1016/0377-2217(93)90075-X

Pyke, D. F., & Cohen, M. A. (1994). Multiproduct integrated production-distribution systems. *European Journal of Operational Research*, *74*(1), 18–49. doi:10.1016/0377-2217(94)90201-1

Raj, P. S., Nichols, W. G., Sterling, J. U., & Moynihan, G. P. (1992). SM2ILE: A heuristic simulation tool for strategic distribution planning. *Computers & Industrial Engineering*, *23*(1-4), 241–244. doi:10.1016/0360-8352(92)90108-V

Ritchie-Dunham, J., Morrice, D. J., Scott, J., & Anderson, E. G. (2000). A strategic supply chain simulation model. Paper presented at the Simulation Conference, 2000, Winter.

Rizk, N., Martel, A., & D'Amours, S. (2006). Multi-item dynamic production-distribution planning in process industries with divergent finishing stages. *Computers & Operations Research*, *33*(12), 3600–3623. doi:10.1016/j.cor.2005.02.047

Ryu, S. W., & Lee, K. K. (2003). A stochastic inventory model of dual sourced supply chain with lead-time reduction. *International Journal of Production Economics*, *81-82*, 513–524. doi:10.1016/S0925-5273(02)00294-3

Safaei, A. S., & Moattar Husseini, S. M., Z.-Farahani, R., Jolai, F., & Ghodsypour, S. H. (2010). Integrated multi-site production-distribution planning in supply chain by hybrid modelling. *International Journal of Production Research*, *48*(14), 4043–4069. doi:10.1080/00207540902791777

Sajadieh, M. S., Jokar, M. R. A., & Modarres, M. (2009). Developing a coordinated vendor-buyer model in two-stage supply chains with stochastic lead-times. *Computers & Operations Research*, *36*(8), 2484–2489. doi:10.1016/j.cor.2008.10.001

Sarimveis, H., Patrinos, P., Tarantilis, C. D., & Kiranoudis, C. T. (2008). Dynamic modeling and control of supply chain systems: A review. *Computers & Operations Research*, *35*(11), 3530–3561. doi:10.1016/j.cor.2007.01.017

Sarimveis, H., Patrinos, P., Tarantilis, C. D., & Kiranoudis, C. T. (2008). Dynamic modeling and control of supply chain systems: A review. *Computers & Operations Research*, *35*(11), 3530–3561. doi:10.1016/j.cor.2007.01.017

Sarker, R., Kara, S., Freeman, G., Kayis, B., Ray, T., & Abbass, H. (2005). A multi-agent simulation study for supply chain operation. Paper presented at the Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce.

Sarmiento, A. M., & Nagi, R. (1999). A review of integrated analysis of production-distribution systems. *IIE Transactions*, *31*(11), 1061–1074. doi:10.1080/07408179908969907

Selim, H., Araz, C., & Ozkarahan, I. (2008). Collaborative production-distribution planning in supply chain: A fuzzy goal programming approach. *Transportation Research Part E, Logistics and Transportation Review*, *44*(3), 396–419. doi:10.1016/j.tre.2006.11.001

Sobhi-Najafabadi, B. (2002). Optimal design of permanent magnet generators. PhD thesis, University of South Australia, Adelaide.

Stadtler, H., & Kilger, C. (2005). *Supply chain management and advanced planning: Concepts, models, software, and case studies*. Berlin, Germany: Springer.

Surie, C., & Wagner, M. (2005). Supply chain analysis. In H. Stadtler & C. Kilger (Eds.), Supply chain management and advanced planning: concepts, models, software, and case studies (3rd Edition ed.). Berlin, Germany: Springer.

Syam, S. S. (2002). A model and methodologies for the location problem with logistical components. *Computers & Operations Research*, *29*(9), 1173–1193. doi:10.1016/S0305-0548(01)00023-5

Syarif, A., Yun, Y., & Gen, M. (2002). Study on multi-stage logistic chain network: A spanning tree-based genetic algorithm approach. *Computers & Industrial Engineering*, *43*(1-2), 299–314. doi:10.1016/S0360-8352(02)00076-1

Tarokh, M. J., & Golkar, M. (2006). Supply chain simulation methods. Paper presented at the IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI '06.

Tasan, A. (2006). A two step approach for the integrated production and distribution planning of a supply chain. *Intelligent Computing, LNCS*, *4113*, 883–888. doi:10.1007/11816157_105

Tavakkoli-Moghaddam, R., Rabbani, M., Gharehgozli, A. H., & Zaerpour, N. (2007). A fuzzy aggregate production planning model for make-to-stock environments. Paper presented at the 2007 IEEE International Conference on Industrial Engineering and Engineering Management. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4419464&isnumber=4419131

Techawiboonwong, A., & Yenradee, P. (2002). Aggregate production planning using Spreadsheet Solver: Model and case study. *ScienceAsia*, *28*, 291–300. doi:10.2306/scienceasia1513-1874.2002.28.291

The-cost-reduction-consultant. (2008). Cost reduction news: Cost control boost for Zara owner. Retrieved 05 October, 2009, from http://www.the-cost-reduction-consultant.com/CostReductionNews.html

Thomas, D. J., & Griffin, P. M. (1996). Coordinated supply chain management. *European Journal of Operational Research*, *94*(1), 1–15. doi:10.1016/0377-2217(96)00098-7

Torabi, S. A., & Hassini, E. (2009). Multi-site production planning integrating procurement and distribution plans in multi-echelon supply chains: An interactive fuzzy goal programming approach. *International Journal of Production Research*, *47*(19), 5475–5499. doi:10.1080/00207540801905460

Tsiakis, P., & Papageorgiou, L. G. (2008). Optimal production allocation and distribution supply chain networks. *International Journal of Production Economics*, *111*(2), 468–483. doi:10.1016/j.ijpe.2007.02.035

Van Donselaar, K., & Sharman, G. (1997). An innovative survey in the transportation and distribution sector. *International Journal of Operations & Production Management*, *17*(7/8), 707–720. doi:10.1108/01443579710175628

Vidal, C. J., & Goetschalckx, M. (1997). Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European Journal of Operational Research*, *98*(1), 1–18. doi:10.1016/S0377-2217(97)80080-X

Wang, D., & Fang, S.-C. (1997). A genetics-based approach for aggregated production planning in a fuzzy environment. IEEE Transactions on Systems, Man and Cybernetics. *Part A*, *27*(5), 636–645.

Wang, R.-C., & Fang, H.-H. (2001). Aggregate production planning with multiple objectives in a fuzzy environment. *European Journal of Operational Research*, *133*(3), 521–536. doi:10.1016/S0377-2217(00)00196-X

Wang, R.-C., & Liang, T.-F. (2004). Application of fuzzy multi-objective linear programming to aggregate production planning. *Computers & Industrial Engineering*, *46*(1), 17–41. doi:10.1016/j.cie.2003.09.009

Williams, E. J., & Gunal, A. (2003). Supply chain simulation and analysis with SIMFLEX. Paper presented at the 2003 Winter Simulation Conference, USA.

Xing, K. (2006). Design for upgradability: Modelling and optimisation. PhD thesis, University of South Australlia, Adelaide.

Xu, J., He, Y., & Gen, M. (2009). A class of random fuzzy programming and its application to supply chain design. *Computers & Industrial Engineering*, *56*(3), 937–950. doi:10.1016/j.cie.2008.09.045

Yan, H., Yu, Z., & Edwin Cheng, T. C. (2003). A strategic model for supply chain design with logical constraints: Formulation and solution. *Computers & Operations Research*, *30*(14), 2135–2155. doi:10.1016/S0305-0548(02)00127-2

Yan, T., Luh, P. B., Weidong, F., & Narimatsu, K. (2003). Supply chain performance evaluation: A simulation study. Paper presented at the IEEE International Conference on Robotics and Automation, ICRA ‘03.

Yeh, W.-C. (2005). A hybrid heuristic algorithm for the multistage supply chain network problem. *International Journal of Advanced Manufacturing Technology*, *26*, 675–685. doi:10.1007/s00170-003-2025-z

Yeh, W.-C. (2006). An efficient memetic algorithm for the multi-stage supply chain network problem. *International Journal of Advanced Manufacturing Technology*, *29*(7), 803–813. doi:10.1007/s00170-005-2556-6

Yilmaz, P., & Çatay, B. (2006). Strategic level three-stage production distribution planning with capacity expansion. *Computers & Industrial Engineering*, *51*(4), 609–620. doi:10.1016/j.cie.2006.05.004

Yimer, A. D., & Demirli, K. (2010). A genetic approach to two-phase optimization of dynamic supply chain scheduling. *Computers & Industrial Engineering*, *58*(3), 411–422. doi:10.1016/j.cie.2009.01.010

This work was previously published in Supply Chain Sustainability and Raw Material Management: Concepts and Processes, edited by Reza Zanjirani Farahani, Shabnam Rezapour and Laleh Kardar, pp. 102-127, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 79
E-Government Clusters:
From Framework to Implementation

Kristian J. Sund
Middlesex University Business School, UK

Ajay Kumar Reddy Adala
Centre for e-Governance, India

ABSTRACT

The concept of industrial clusters has received much attention in the literature over the past few decades and many examples of clusters exist today in a variety of industries, from manufacturing to services. Within such clusters, competitive cost and innovation advantages are generated through co-location. Very recently several examples of e-government clusters have emerged. This chapter offers a conceptualization of what an e-government cluster is, and how it may be different from other industrial clusters. This chapter is an attempt to formulate a framework for e-government clusters and bring out the necessary conditions for policy decisions to support the creation of such a cluster. An attempt has also been made to validate the proposed framework on the basis of case studies and to derive some recommendations to sustain the operation of e-government clusters.

INTRODUCTION

Information and Communication Technologies (ICTs) are increasingly being used in government as indispensable tools, in an effort to transform them into efficient, effective, transparent and reliable organizations at national, regional and local levels. Electronic government (e-Government) has been defined as the application of ICTs in transforming the internal and external relationships of governments [UN World Public Sector Report 2003]. While many governments across the world have implemented e-government projects, many have failed to live up to expectations due to the design-reality gaps, which were often not considered during the planning stages (Heeks, 2001).

Since e-government is as much about government as about electronic ('e') (Riley, 2002),

DOI: 10.4018/978-1-4666-1945-6.ch079

its implementation is likely to involve active and shared participation of several players from private, public and social sectors. Often, governments face a shortage of the right expertise, right technologies and right solutions to successfully implement e-Government projects. This situation is similar to that faced by most industries during their initial stages, especially when they set-up in new locations or regions. Subsequently, the availability of right skills and technologies improves with several firms joining the region resulting in the formation of clusters.

The concept of industrial clusters has received much attention in the literature over the past few decades and many examples of clusters exist today in a variety of industries, from manufacturing to services (Porter, 1990). Recently, several projects have seen the light of day under the heading of "e-Government clusters". Examples of these include the e-Government cluster in Hungary, Ubiquitous IT (u-IT) cluster in Korea and the GAUDI e-Government cluster of the European Union. The e-Government Cluster in Hungary has been the initiative of the private sector with active support from the Government of Hungary at local, regional and national levels. The u-IT cluster of Korea is driven by the government and primarily focuses on the development of ubiquitous technologies for the implementation of e-Government at the local level. The GAUDI e-Government cluster is the result of the PRELUDE project initiated by the European Commission under the 5th Framework Program with an aim to establish 9 European Clusters for Innovations (ECI). The GAUDI e-Government cluster is a consolidation of the Lombardy cluster of Italy, Kouvala cluster of Finland and Catalonia cluster of Spain. Unfortunately, although the concept of industrial clusters has been widely examined in the literature, no framework has yet been developed within the context of e-Government, something from which e-government clusters such as those mentioned previously might benefit. This chapter proposes to develop such a framework answering the basic question '*What are the necessary pre-conditions for Government to support the creation of e-Government cluster?*' As a logical sequence to this question, an attempt has been made to address the question '*What can be the likely success factors which sustain the operation of e-Government cluster?*'.

Historically, different regional economies around the world had within them specialized industries operating in the form of so-called clusters. These clusters were composed of main firms, which produce industry output, and supporting firms which actively contribute to the efficient operation of the main firms. Through a mix of competition and cooperation, such firms tend to create sustainable competitive advantages within their industry and the cluster, as compared to other isolated players of the industry (Porter, 1990). Famous examples of such regional clusters include Silicon Valley of the United States, the pharmaceutical cluster centered on Basel in Switzerland, and fashion-leather cluster of Northern Italy. Such agglomeration of firms brings about improvement and innovation owing to the mutual reinforcement of cluster-based factors, subsequently leading to sustained competitive advantage. In other words, firms within such a cluster benefit from agglomeration economies with reinforced effects. The cluster becomes a vehicle for maintaining diversity and overcoming the inward focus, inertia and inflexibility that slows or blocks competitive upgrading and new entry (Porter, 1990).

It is with these advantages in mind that attempt have been made around the world to create e-Government clusters. However, it remains unclear whether the concept of cluster can be applied in the context of e-Government. Neither can one readily find a specific framework for e-Government clusters in the literature. Simply transposing the original notion of clusters, as defined by economic geographers, to the field of e-Government may not be the best way to plan or enhance such a cluster.

The most common formulation of industry clusters is probably that of Prof. Michael Porter of Harvard Business School, made famous by his 1990 book, “The Competitive Advantage of Nations”. In this chapter we adopt a similar approach, and analyze the Porter “diamond” model of clusters from the perspective of its application to the e-Government “sector”. The e-Government clusters in Hungary, Korea and EU have been critically examined through interviews, e-mail enquiries and secondary research.

The outcome of this chapter is a detailed framework on e-Government Cluster, adapted from the “diamond” model of Michael Porter, and a description of various factors to be considered when implementing e-Government clusters. Also, the importance of each factor and their role in e-Government competitiveness has been assessed. This chapter also provides insights on critical aspects of clusters which governments should carefully examine prior to its decision on supporting the creation of e-Government cluster. The validity and necessity of physical proximity of firms while forming such e-Government clusters has also been considered, given the possibilities for remote working offered by modern ICTs. The results generated not only would guide e-Government cluster investments, but also more generally the outsourcing of e-Government projects.

CLUSTERS AND e-GOVERNMENT

Clusters

Defining Clusters

Clustering, in broad generic terms, involves the grouping together of tasks or entities with similar goals and attributes. The subject of clusters is a traditional concept which is increasingly being used currently in diverse fields and areas to achieve goals in a collaborative and cooperative yet competitive environment. For example, a marketing cluster involves the segmentation of customer market based on the nature of the product, age of the customer, income level of the customer etc. Similarly, a research cluster involves the grouping of teams working on a particular subject, technology, device etc. Clusters, in general, may be categorized as marketing clusters, academic / research clusters, partnership / alliance clusters, industrial clusters, project clusters, knowledge clusters etc. With the rapid growth of information and communication technologies, online partnership clusters have seen the day in the form of online communities and online forums.

Today, the concept of industrial clusters is well known in management communities owing to the popular “diamond” cluster theory, published by Professor Michael Porter, explaining the reasons for growth or decline of competitiveness of a nation or region. The concept of industrial clusters has become so widely known in the management, research and industrial sector that the word “cluster” is quite often directly associated with the concept of industrial cluster. Prominent examples of industrial clusters include ICT cluster in Silicon Valley (USA), fashion leather cluster in Italy, wine cluster in California (USA), the forestry cluster in Finland and consumer electronics cluster in Japan.

The Formation of Clusters

It is very interesting to study the formation of clusters in various parts of the world. The forest cluster of Finland was more of a natural phenomenon owing to the ready availability of abundant and economically accessible natural resources (Rouvinen & Ylä-Anttila, 1999). The fashion leather cluster of Italy has been the result of the continued and sustained family business inheritance, whilst the consumer electronic cluster of Japan has been due more to the specific needs and wants of its citizens. Studies have shown that a strong home demand with favorable factor conditions positively influence the development of demand-related and input-oriented industries in a region (Porter, 1990). Furthermore, competitive advantage in the form of location becomes an important factor in determining the survival or exit of individual firms over the course of the industry life cycle (Sund, 2006).

The sustained operation of firms within clusters are said to induce the development of other firms operating in related and supporting industries, supporting the main firms producing the industry outputs. The intense rivalry among organizations in a cluster often leads to stiff competition forcing companies to enhance their product features and quality, through process improvement, technology upgrades or innovative solution deployment. The firms within the cluster tend to sub-contract their procurement and other activities to the related and supporting firms set-up within the cluster as they would be able to provide more efficient and reliable services. This would further attract firms aiming for market share in the demand for suppliers generated by the cluster.

The nature of firm strategy and structure is likely to have an influence on the competitiveness of firms within the cluster, leading to improved factor conditions and demand conditions. The companies not attempting to innovate, improve and upgrade, would gradually lose ground and disappear from the cluster.

The phenomenon of the formation of clusters is well explained by Michael Porter in his "Diamond" theory, which has the following four determinants.

- Factor conditions
- Demand conditions
- Firm strategy, structure and rivalry
- Related & Supporting Industries

The reinforcement among the determinants results in the cluster being competitive, thereby, enhancing the productivity and economic potential of the region. The Porter "diamond" is illustrated in Figure 1.

Factor Conditions

The factor conditions are those essential inputs accounting for the successful operation of an industry. They drive the economy of the industry and the growth of the region. They include natural resources, demographic conditions, location geography, production facilities, investment and capital potential, the presence of highly skilled and quality human resources, research infrastructure, universities present in the region, and so forth.

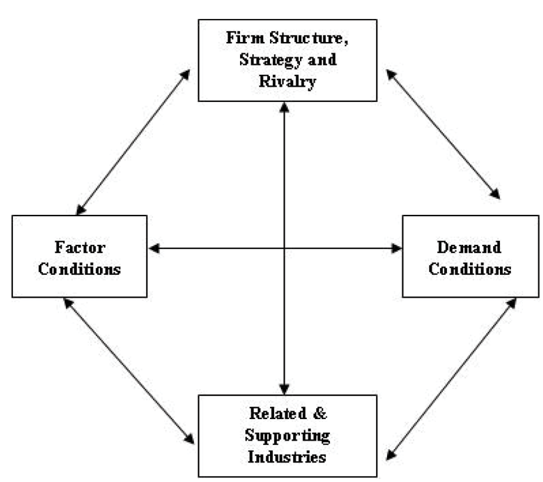
Demand Conditions

A strong home demand is usually considered a good driver to create competitiveness within the cluster and the region. The demand may be local, regional, national or even international. A strong home demand is not necessarily a question of quantity, but largely a question of the quality demanded. More demanding customers lead to better products that ultimately will be more competitive in the global market place. Thus, a sophisticated local demand is usually considered as a take-off point to stimulate the formation of clusters (Rychen & Zimmerman, 2008).

Related and Supporting Industries

Firms show an increased tendency to purchase from local suppliers if no compromise is seen in

Figure 1. Porter 'Diamond' Cluster Framework



quality and reliability of products and services when compared to purchases from established far-off suppliers. This would induce suppliers, which are specialized in a certain field, in setting-up their base within the cluster region. Subsequently, with time, these related and supporting firms tend to become a part of the overall value chain, enhancing the drive to innovate, improve and upgrade, strengthening their competitive position.

Firm Structure, Strategy and Rivalry

The attractiveness of an industry may not be reliably indicated by size, rapid growth or newness of technology, but rather by industry structure. The mechanism by which companies are created, organized and managed together with the approaches adopted by them to handle firm rivalry are known to have far-reaching and influential effect on the competitiveness of the cluster and, thereby, the economic potential of the region. It has been seen that healthy firm rivalry results in a local race to introduce new technologies, enhance productivity and lower costs.

Role of 'Chance' and 'Government'

The four determinants of the cluster as illustrated above are considered by Porter as internal sources of competitiveness. He further identified the external sources of cluster competitiveness as chance and the role of government. The external sources have a potential capability to change the dynamics of cluster and, thus, reorient it in different directions.

Chance

Chance events may be considered as occurrences with less accountability to the circumstances in a nation and are often largely outside the power of firms (and often national government) to influence. Possible examples of chance are acts of nature, major technological breakthrough, drastic changes in input costs such as an oil crisis, sudden shifts in financial markets, fluctuation in national or global demand, offset of war and so forth. Innovative and sustainable leadership often becomes a key word to drive competitiveness of the cluster in cases of such chance events.

Government

Government is quite often seen as a catalyst in the case of clusters. In some instances, governments become major buyers of the products produced in a particular cluster, for example in the case of defense goods, telecommunication equipment or aircraft. More generally though, the government is seen as a facilitator to effectively support various firms in their operations with a view to enhance competitiveness of the cluster and the region. Government policies on product standards, regulations, tax, anti-trust laws, education, and so forth, are seen to greatly influence the main forces of firm structure, strategy and rivalry, demand conditions, factor conditions and the health of related and supporting industries.

Cluster Benefits

The benefits associated with clusters affect competition in three broad ways – increasing productivity, driving the pace of innovation and stimulating the creation of new businesses. A cluster allows each member to benefit as if it had greater scale or as if it had joined with others formally – without requiring it to sacrifice its flexibility. In other words, it can be said a cluster is more than just the sum of its parts. The benefits of such a cluster are listed as follows.

e-Government

Defining e-Government

The concept of electronic Government (e-Government) has been defined in several ways by various authors and organizations. Some of them are as follows.

Information for Development Program (InfoDev 2002)

e-Government is the use of information and communication technologies to transform government by making it more accessible, effective and accountable

Organisation for Economic Cooperation and Development (www.oecd.org)

e-Government is defined as a capacity to transform public administration through the use of ICTs

World Bank (www.web.worldbank.org)

e-Government refers to the use by government agencies of information technologies (such as Wide Area Networks, Internet, Mobile Computing) that have the ability to transform relationships with citizens, businesses and other arms of government

European Union Commission (www.eu.europa.eu/egovernment)

e-Government is the use of ICTs in public administration combined with organizational change and new skills in order to improve public services and democratic services and strengthen support for public policies

United Nations (www.unpan.org, 'World Public Sector Report 2003')

e-Government is a government that applies ICTs to transform its internal and external relationships

Gartner Research (Gartner Group, www.gartner.com)

e-Government is the continuous optimization of government service delivery, constituency participation and governance by transforming internal and external relationships through technology, the internet and new media

Looking at the various definitions, the definition of e-Government seems to revolve around three important issues which are as follows.

1. Implementation of ICTs
2. Organizational change
3. Transformation in internal and external relationships

E-Government is thus typically seen as having a potential to optimize government processes, making them more effective and efficient. If e-Government is seen only to implement ICTs, it would be seen as an old car fitted with a new engine resulting in an inefficient and uneconomical drive. More than just the implementation of ICTs, e-Government involves the reform of government policies, relationships with various State and non-State actors, the re-engineering of operational processes, and innovation in service delivery mechanisms. Thus, the innovation involved has the power to change the relationship between levels of government and within government departments (Alberti & Bertucci 2006). ICTs need to be considered as tools to facilitate online transactions and achieve the broader goals of the government (Titah and Barki 2005). It is of course worth underlining the role of ICTs as a tool for development and not as a goal in itself (Misuraca, 2006).

e-Government Frameworks

E-Government frameworks have been designed in different ways by various organizations and institutions. Since countries vary in their political mandates, traditions, cultures, legacy systems, government working norms etc., it is a well accepted fact that one single e-Government framework may not be deemed fit universally.

This section contains some condensed examples of e-Government frameworks.

GePeGi (Korea) e-Government Framework

The e-Government framework designed by the Global e-Policy e-Government Institute (GePeGi), SungKyunKwan University, Seoul, Korea comprises of three factors: demand factors, supply factors and policy factors (Korea e-Government 2007). The supply factors are related to the development of ICTs and include network infrastructure, hardware infrastructure and application infrastructure. The demand factors are related to the maturity of civil society and include information management, e-capacity and economic support. The policy factors are related to the maturity of government and include IT policy, Institutional and Legal support policy and IT manpower policy.

NISG (India) e-Government Framework

The National Institute for Smart Government (NISG), India, is of the view that the full potential of e-Government can be realized only through a focused attention on people, process, technology and resources appropriately (Satyanaryana, J 2004), which are viewed as the four pillars of e-Government framework. The people pillar is related to the human element involved in e-Government and includes training, capacity building, change management, institution building, organization structure, governance structure, consultancy, monitoring, evaluation, awareness and communication. The process pillar is related to government workflows which need to be re-examined and includes process design and development, administrative reforms, process re-engineering, program / project management, procurement, and service levels. The technology pillar is related to the deployment of technologies in government and includes enterprise architecture, information infrastructure, standards and interoperability, security and privacy policies, identity management, biometrics, private key infrastructure and digital signatures. The resources pillar relates to the knowledge base on e-Government and includes public private partnership frameworks, service level agreement frameworks, business case analysis, mission teams, knowledge management, and user charges on services.

Singapore e-Government Framework

The e-Government framework designed for Singapore is based on different government interfaces with citizens (G2C), employees (G2G) and businesses (G2B). While the G2C and G2B services involve electronic service delivery as the main component, the G2G services are seen to involve knowledge management and education in information and communications (Altameem, T., Zairi, M. and Alshawi, S. 2006). The framework also suggests a strong focus on ICT infrastructure and operational efficiency in order to ensure a seamless interface and interoperability in the government interactions and transactions with employees, citizens and businesses.

Common to these three e-Government frameworks, it appears that the successful implementation of e-Government is likely to involve active collaboration among various e-Government actors, including political organizations, the public sector, business organizations, trade associations, training and research organizations, citizens and civil societies, deploying right resources at the right time and promoting productivity and innovation in government.

e-Government Success Factors

Many e-Government projects tend to fail owing to popular misconception of ICTs being ends rather than means (West 2007). It is important to realize the role of ICTs prior to any conceptualization of e-Government. ICTs are just a tool to provide an enabling environment to enhance efficiency and promote innovation (Misuraca, 2007). Several authors and organizations have suggested factors which account for the successful implementation of e-Government, some of which are as follows:

- Richard Heeks (2001): The success of e-Governance depends on building e-readiness and the closure of gaps (design-reality gaps, private-public gaps, country-context gaps).
- InfoDev (2002): Success factors for e-Government include process reform, leadership, strategic investment, collaboration and civic engagement.
- Jorg Becker (2004): The potential e-Government success factors can be viewed as organizational responsibility for Government, e-Government awareness, budgetary funding and organizational change.
- Altameem (2006): The critical success factors for e-Government can be categorized into governing factors, technical factors and organizational factors.
- Mila Gasco (2008): The key factors for the adoption of ICTs in government are political and personal leadership, strategic alignment between political strategies and IT initiatives, deployment of the encompassing institutionalization strategies, Information Technology to be a part of a more comprehensive strategy, and the management and architecture of corporate information resources.

DEFINING THE e-GOVERNMENT CLUSTER

An e-Government cluster may be defined as a '*group of inter-related and competing firms and organizations that offer products and services in the broad area of e-Government, and derive synergistic advantages from their co-location in a given region*'.

Typically, firms involved in e-Government services deliver hardware or software solutions to governments and government organizations with a view to enhance their efficiency, effectiveness, transparency, accountability and reliability (Mahapatra and Perumal 2006). Services range from the identification of system requirements, which is a pre-designing phase involving interactions and discussions with government and other stakeholders in order to identify the precise needs and requirements, all the way to servicing and advising government on the use of existing systems.

As e-Government is fast attaining the status of an industry sector in its own right, e-Government clusters can be thought of along similar lines as any other industry cluster, i.e. as a group of main firms operating in e-Government related software services and a group of supporting firms which operate in hardware, network, consultancy, training and research. As a necessary condition for an e-Government cluster to exist, strong demand from customers, in the form of government offices or agencies, need to be in close vicinity of the cluster.

The implementation of e-Government has been gaining momentum across the globe. Owing perhaps to a shortage of e-Government expertise, as compared to other IT expertise, it is often a difficult exercise for governments to identify the right partner who can conceptualize, initiate and implement e-Government projects for them. The choice of partner is often strategic in nature, and governments are sometimes seen to favor local partners. The common practice of issuing tenders or Requests-for-Proposals (RFPs), either nationally or globally, in order to select the right partner, is at times both costly and cumbersome, involving complicated pre-qualification criteria, technical criteria and commercial quote criteria. The selection of a wrong partner in the process is seen as a potential risk, in which case governments most often have to restart the whole process once again (Jaeger, Paul, Thomson and Kim 2003). It is perhaps with this in mind that efforts have been made in certain regions to build regional competence in e-Government solutions, through the creation of clusters.

A PROPOSED FRAMEWORK

Having defined the e-Government cluster and realizing the need for productivity and innovation in government, we hereby attempt to propose a framework which can serve as a basis for understanding such clusters. The proposed framework is based on Porter "diamond" framework, which was substantiated by several clusters across various countries in the world.

Since the Porter "diamond" framework is based on manufacturing industries like hardware, wine, chemical, fashion leather, consumer electronic etc., producing ready-to-use products, an attempt has been made to extend the "diamond" framework to illustrate the dynamics of e-Government cluster. Given the specificity of e-Government as a sector, the Porter "diamond" has been extended by including more constituents with a view to present a comprehensive framework for a specialized sector like e-Government.

As it is believed that favorable factor conditions and demand conditions are the likely initial drivers to attract organizations to establish themselves in a region, two additional "diamonds", the factor diamond and demand diamond, have been proposed, which would facilitate in creating favorable factor conditions and demand conditions. The factor diamond and demand diamond are based on networking dynamics among various e-Government actors. The proposed e-Government cluster framework is therefore a combination of three diamonds, the main diamond, the demand diamond and the factor diamond. The main diamond is the original "diamond" as suggested by Porter and serves as a parent framework for the other two diamonds.

The demand diamond is a framework involving various actors, whose active collaboration and cooperation is likely to generate favorable e-Government demand conditions. Unlike other industrial clusters where demand is mostly driven by the end user, the demand for e-Government is initially driven by the Government and its various agencies and organizations.

The factor diamond is a framework involving various actors, whose relationship with each other is likely to create favorable e-Government factor conditions, Unlike other industrial clusters where factor conditions are more focused on the availability of raw materials, transport infrastructure and so forth, the factor conditions for an e-Government cluster specifically include the availability of skilled manpower, network and broadband services, software solutions, hardware services and other related consultancy services.

The proposed framework is illustrated in Figure 2.

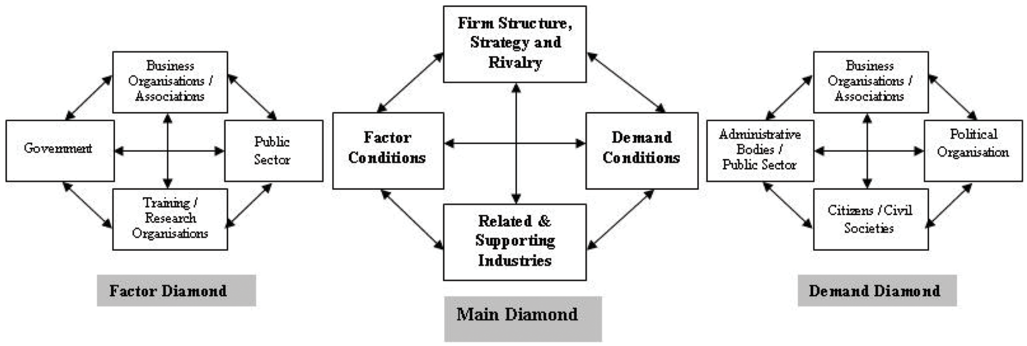
Main Diamond

Demand Conditions

The demand for e-Government services is seen as a necessary condition to enable the creation of an e-Government cluster in a region. Some of the factors which facilitate in generating favorable demand conditions are as follows.

1. Commitment to the implementation of e-Government in all government departments, agencies and organizations.
2. Adherence to e-Government strategies to achieve e-Government vision and objectives.
3. Intention to deploy high quality and innovative e-Government solutions with strong focus on reliability and sustainability.
4. Focus to improve democracy through provision of technology to citizens.
5. Implementation of citizen participatory mechanisms through online feedback system, discussion forums, citizen surveys, messaging applications etc.

Figure 2. Proposed e-Government Framework (based on Porter (1990))



6. Strong political will and leadership to transform labor-intensive methodology to technology-intensive methodology.
7. Partnerships / relationships with other State and non-State actors in order to utilize their knowledge, skills and expertise.
8. Right attitude and mindset among staff and officials at all levels of the government.

Factor Conditions

The factor conditions in a location / region may be categorized into basic factors and advanced factors. The basic factor conditions include availability of raw materials, power, land etc., and the advanced factor conditions include availability of expertise, skills, data storage centers etc. Both these factors need to be critically examined prior to setting-up of e-Government cluster. Some of the aspects which result in creating favorable factor conditions are as follows.

1. Easy accessibility of the region by various modes of transport.
2. Business-friendly environment at all levels of government.
3. Presence of strong telecommunication, broadband and network infrastructure ensuring reliable communication services.
4. Availability of skilled manpower (mainly engineers) in the region.
5. Willingness of skilled personnel to work in the region.
6. Presence of reasonably priced building infrastructure with basic facilities including power, security and broadband connectivity.
7. Possible presence of high capacity server infrastructure and data storage centers.

Related and Supporting Industries

The related and supporting industries like IT hardware, telecommunication and so forth, are likely to impact the sustenance of an e-Government cluster as these firms provide value-added services to the main firms of the cluster. A major driver attracting and retaining the supporting industries would be the strong home demand for e-Government services resulting from a strong commitment to implement e-Government strategies. Firms specialized in hardware, telecom, network infrastructure, training, academics, research, knowledge management and consultancy would add value to the overall value chain of the e-Government cluster.

Firm Structure, Strategy and Rivalry

The firm structure and strategy are often key areas to incorporate necessary changes needed to strengthen competitive position in e-Government market. Also, healthy firm rivalry is seen as an important mechanism to develop the attitude of continual improvement in delivering e-Government services. However, the presence of firm rivalry may not be seen in prominence during the initial stages of e-Government cluster formation. Subsequently, with the establishment of the e-Government cluster, firm rivalry would be seen as one of the prime factors enhancing the productivity and competitiveness among cluster members.

Demand Diamond

The Demand Diamond involves active relationships among four key actors of e-Government. These are the administrative bodies and public sector in general, political organizations, business bodies and trade associations and citizens and civil societies. These determinant actors and the nature of their relationship are likely to help generate sustained demand for e-Government services. Also, these relationships are likely to transform traditional-based governments into network-based governments utilizing the right skills and expertise, readily available within the network. The Demand Diamond in Figure 3 is therefore to be seen as a map of the key actors involved in forming demand for e-Government, and we would expect to see regions with strong collaboration between these actors as being more receptive to e-Government.

Factor Diamond

The advanced factor conditions needed for setting-up an e-Government cluster, apart from the basic factor conditions, may be categorized broadly as: ICT infrastructure, Capital infrastructure, Skilled personnel and Knowledge. The ICT infrastructure includes network infrastructure, hardware infrastructure, broadband connectivity, software facilities, data storage infrastructure, security infrastructure and other supporting components. The ICT infrastructure would be one of the influential factors attracting firms to the e-Government cluster owing to its capability in serving as a backbone supporting numerous IT-based solutions and services. The lack of basic ICT infrastructure is often seen as a hindrance in the realization of e-Government (Jaeger & Thomson, 2003).

The factor diamond of the proposed framework involves relationships among various actors involved in creating favorable factor conditions. The factor diamond is illustrated in Figure 4.

The Government through regular investment and through its e-Government projects can deploy elements like server infrastructure, network infrastructure, routing equipment and other related accessories, substantially improving the ICT infrastructure of the region (Chowdhury, Habib, Wahidul and Kushchu 2006). Likewise, the public sector contributes to the e-Government factor conditions through their strategies to implement IT solutions in their organizations. The presence

Figure 3. Demand Diamond

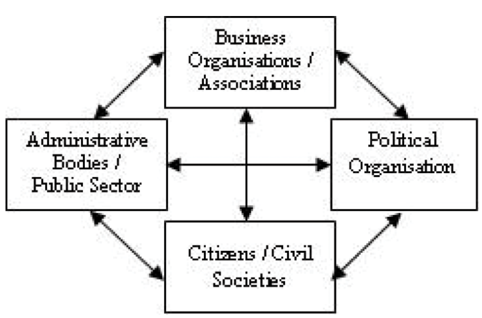
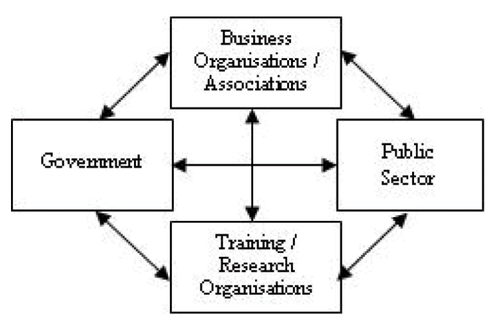


Figure 4. Factor Diamond



of business organizations and trade associations in the region may improve basic factor conditions and some of the advanced factor conditions, which include the generation of skilled personnel and creation of knowledge societies. Training and research organizations are of course instrumental in generating skilled personnel through education, work-related training programs and research projects.

CASE STUDIES

In order to examine the proposed e-Government Cluster framework, three case studies of e-Government clusters are examined: the Korean u-IT Cluster, the Hungarian e-Government Cluster and the GAUDI e-Government Cluster. The analysis is based mostly on the secondary research data collected from various sources, although some critical data was obtained through telephone interviews with concerned authorities and officials. A brief description of each cluster is as follows.

u-IT Cluster, Seoul, South Korea

The u-IT cluster plan has been framed for fostering industries of RFID and USN, which are the core technologies for construction of u-City and the next-generation growth drivers as clearly articulated in Korea's IT839 strategy. The u-IT cluster is planned to form an innovative cluster by attracting related leading companies and R&D centers. Furthermore, it will build a globally-competitive cluster that can continue to grow by facilitating technological innovations in the RFID and USN industries and by creating new businesses for different value chains and contribute to the formation of the IT hub in Northeast Asia. The main purpose of this master-planned cluster, which is being developed on 1,500 acres of reclaimed land along Incheon's waterfront, is to give opportunities to many leading RFID / USN companies around the world to complete RFID / USN products in the most efficient and cost-saving way.

The Ministry of Knowledge Economy and the Incheon Metropolitan City are in charge of the project and the cluster will be formed in the knowledge and information industry complex of Songdo area. They have made investments totaling $313.7 million to establish the "World's Best Infrastructure Facility for RFID / USN technologies" with the size of the building site amounting to 81,275 sq. mtrs. in the Incheon Free Economic Zone of Songdo from 2006 to 2010. The relevant public sectors shared the view that it was important to create a world-renowned RFID / USN industrial collective complex, encompassing research, production and services. The country believes that by establishing the u-IT cluster center, the government has sought to bridge the technological gap between Korea and other technologically advanced countries (Korea IT Times 2008).

Hungarian e-Government Cluster

The first Hungarian e-Government Cluster was founded in June 2007. It was a unique initiative and the only cluster Hungary has in e-Government, though plans are underway to create regional clusters in future. The e-Government Cluster aims to create an effective e-Government as well as to build a market by common enforcement of interests and harmonizing activities. The Cluster covers a wide spectrum with comprehensive specific competencies assuring the interaction of different views.

The cluster members include advisor companies, consultancies, universities, system integrator companies, knowledge centers, local government and alliances of towns. While members can function independently, their ability to develop synergistic relationships and cooperate in a network has enabled them to reap the benefits of working together. However, members' first priority has been to improve services significantly across the whole field of local government.

The managing agent of the Hungarian e-Government cluster is Albacomp Zrt., which is well known for its major network and system integration projects apart from operating Hungary's largest PC and notebook assembly plant. The company has been actively involved in several IT projects, some in part funded by the Hungarian government, and has played a significant role as an integrated IT solutions provider to local governments that are in the process of setting up their local e-Government services.

The cluster performed at least half a dozen projects and prepared proposals for more than a dozen projects. The cluster is looking forward to enter into partnerships with international firms and organizations to improve the competitiveness and productivity. The cluster will apply for an audit process to achieve "Accredited Polus Cluster" qualification. In the next 5-10 years, the cluster would be in a position to market itself on Pan-European level and possibly global level.

GAUDI e-Government Cluster: European Cluster for Innovation

The GAUDI e-Government cluster, which was a result of the PRELUDE project initiated by the European Commission under 5th Framework Programme with an aim to establish 9 European Clusters for Innovations (ECI), provided a critical bridge to the G2C aspects suggested by GUARANTEE and the G2B aspects of the digital economy being driven by EDEn (Zalisova 2006).

The GAUDI cluster was a consolidation of three regional clusters: Kouvola (Kymenlaakso) of Finland, Lombardy of Italy and Catalonia of Spain. The members of each regional cluster are shown in Table 1.

The activities of the cluster included the preparation of an e-Government Roadmap, e-Transaction projects, organizing the Internet Global Congress and facilitating the exchange of experiences, good practices, know-how and demonstration projects.

ANALYSIS

The main 'Diamond' as illustrated in the proposed e-Government Cluster Framework had been firmly established by Porter based on industrial clusters across various countries of the world. Since e-Government is fast attaining the status of industry sector, bearing most of the industry characteristics, the 'main diamond' of the proposed e-Government cluster framework has not been analyzed. The differentiating feature between e-Government clusters and industrial clusters is the role of demand diamond to generate demand conditions and the role of factor diamond to create favorable factor conditions. The demand diamond determinants like administrative bodies,

Table 1. Regional Clusters of GAUDI Cluster

Kouvala Cluster	Lombardy Cluster	Catalonia Cluster
• IT companies • National Land Survey • Population Register Centre • Helsinki University of Technology • Lappeenranta University of Technology	• 4 Lombardy Region General Directorate • 5 Public sector special agencies • CEFRIEL (Politecnico Milano) • 5 municipalities and provinces associations • 14 professional associations and enterprises associations • Lombardy Chambers of Commerce Union • 4 trade-unions	• Consorci Localret - 800 municipalities - Catalan association of municipalities counties - Federation of municipalities - 4 Catalan provinces • Generalitat de Catalunya • 4 provincial public administration bodies • 11 Catalan municipalities • 5 universities: UoC, UPC, UB, UPF, UdG • Cat 365

political organizations and civil societies may not be involved for industrial clusters but are likely to have an impact for e-Government clusters. Similarly, the factor diamond determinants like the government, public sector and training organizations may not be influential for industrial clusters but are likely to have substantial contribution for e-Government clusters.

The demand diamond and factor diamond have been analyzed in perspective of the three e-Government cluster cases elaborated in the previous chapter. The Demand Diamond Analysis and Factor Diamond Analysis is an attempt to examine the proposed e-Government cluster framework in the light of real-world examples, in particular its applicability for the creation and sustenance of e-Government clusters.

It needs to be mentioned here that the relevant data on the three clusters was collected through the following mechanisms:

- Websites, publications and journals
- Interactions with concerned officials during the modules of the Executive Master in e-Governance, EPFL
- e-Mail enquiries
- Telephonic interviews

Every attempt has been made to collect as much relevant information as possible so as to achieve near-real analysis.

Demand Diamond Analysis

The demand diamond analysis, which has been based on the level of involvement exhibited by each demand diamond determinant in the generation of demand for e-Government, for each of the clusters, is as follows.

u-IT Cluster, South Korea

The ubiquitous IT Cluster at Incheon is driven by strong demand generated by the Vision 2030 of Korea, which has e-Government as one of the core driving forces, along with the u-IT strategy, IT839 strategy and u-Korea strategy. The u-IT Cluster is a key initiative of the Ministry of Knowledge Economy and the Incheon Metropolitan Authority, with relevant public sectors sharing a common view. It can be derived that the political organizations, administrative bodies and public sector had a very high level of involvement in generating demand for ubiquitous services, which in-turn are directly linked to the generation of demand for e-Government services.

Korea had been implementing ICTs as early as 1980s with the active involvement of its ICT industries. The government was instrumental in establishing basic infrastructure through 11 key e-Government projects during 2002-03 and 31 e-Government Roadmap projects during 2003-07. It is likely that the successful implementation of e-Government projects, due to the active involvement of ICT industries, associations and other research organizations, would have stimulated government to visualize achieving a ubiquitous society. Business organizations had a high involvement while generating demand for future ubiquitous technologies like RFID and USN.

The role of citizens and civil societies in generating the demand for ubiquitous services seems unclear, indicating a possible low level of involvement, except for the mention in the Vision 2030 document to understand customer's needs and offer quality services in a timely manner.

e-Government Cluster, Hungary

The Government of Hungary published a document on 'Information Society' as early as 1999 and issued the first version of the National Information Society Strategy in 2001, which had the implementation of e-Government as one of the priority areas. The Hungarian Parliament launched the e-Parliament programme to support the modernization of Parliament. Also, the Hungarian Parliament adopted the Act of Freedom of

Information by electronic means in 2005, with a goal to establish the legal environment required to create a transparent Digital State. The Hungarian Government formulated the e-Administration Development Strategy 2007-13 which contains the e-Public Administration 2010 strategy. These series of events indicate a high level of involvement from the government (administrative bodies and public sector) in generating demand for e-Government services. Also, since the politicians from various political organizations approved the formulation of e-Parliament and other ICT projects, it shows that the political organizations had considerable involvement in initiating e-Government projects.

Hungary had been implementing e-Government projects since the year 2002 with the help of local and national firms. Due to insufficient domain knowledge within government, they often ended up purchasing expensive high capacity servers and computing equipments beyond their actual needs. The government seemed to greatly benefit from the e-Government cluster as it reduced unnecessary expenditure and enabled it to have a better view of the new system. The involvement of citizens and civil societies in generating demand for e-Government seems to be unclear, indicating a possible low level of involvement, as it was found unmentioned both in secondary research and telephonic interviews.

GAUDI e-Government Cluster

The GAUDI Cluster is a consolidation of three regional clusters (Lombardy Cluster of Italy, Kouvala Cluster of Finland and Catalonia Cluster of Spain) with IT companies, administrative bodies, universities, public sector, business associations, trade unions and other government agencies as members. Unlike the previous two clusters, the GAUDI cluster operates on partnerships among the regional clusters. The physical presence of an organization within a region was not treated as a necessary condition. The cluster had a focused objective to support, promote, contribute and lead e-Government in the European Arena, thus generating demand for e-Government. The administrative bodies and public sector are seen to have a high level of involvement in generating demand for e-Government services as they had prepared an e-Government Roadmap and submitted e-Government proposals to the European Union. The involvement of political organizations seems to be less as there was hardly any mention of such organizations, either in the literature or during telephone interviews.

The involvement of citizens and civil societies was not clearly defined while framing the objectives of the GAUDI cluster, indicating a possible low level of involvement. The only noticeable mention was in providing user-friendly personalized services and multi-channel access using advanced ICTs such as voice-enabled front-end public platforms. However, the need for citizen and civil society involvement was suggested for the 7th Framework Programme of the European Union.

The summary of the Demand Diamond Analysis is shown in Table 2, which gives the level of involvement by demand diamond determinants on each cluster.

From these case studies, it appears that citizen and civil society involvement was not necessarily a systematic feature in the planning of these e-Government clusters. This does not mean that the clusters, and the services and other results emerging from these, are not citizen-centric in themselves, but it does question our proposed framework in terms of what role citizens can realistically play in such clusters.

Factor Diamond Analysis

The factor diamond analysis below is based on the level of influence exhibited by each factor diamond determinant in creating favorable e-Government factor conditions, for each of the clusters.

Table 2. Demand Diamond Analysis

Determinants	u-IT Cluster South Korea	e-Govt. Cluster Hungary	GAUDI Cluster European Union
Political Organisations	High	High	Low
Administrative Bodies / Public Sector	High	High	High
Business Organisations / Associations	High	High	Medium
Citizens / Civil Societies	Low	Low	Low

u-IT Cluster, South Korea

The Korean government has established the basic IT infrastructure through its 11 key e-Government projects (2002-03), which in-turn helped the government to implement 31 e-Government roadmap projects. The government has planned to develop u-City infrastructure in major cities by putting in place 3 infrastructures viz., Broadband Convergence Network (BCN), Ubiquitous Sensor Network (USN) and Next-Generation Internet Protocol (IPv6). From the above, it can be deduced that the government played an important role in building the basic infrastructure. Since the State owned companies in Korea obtain direct contracts from the government, they are also believed to have a substantial role in providing IT infrastructure. Thus, the government and public sector are seen to have a high influence in creating factor conditions.

The premises to host firms in the u-IT cluster have been built by the government with rental costs reported as quite low. Facilities include convenience facilities like dormitories, dining rooms, fitness centers etc. This shows that the government and the public sector have been instrumental in providing adequate and reasonably priced capital infrastructure.

Korea reported in 200X that 77% of population had access to high-speed internet and 88% of population having mobiles. The presence of large corporates and SMEs in and around the cluster region has led to an increase in the skills and expertise of the people. It can be deduced that the business organizations have had a high influence on e-Government factor conditions.

Incheon is an hour drive from Seoul, which is host to numerous universities, training institutions and research & development organizations. These institutions have been operating since a long time and have proven to be a source of rich knowledge and experience. This aspect of readily available knowledge has induced firms to set-up their base in the u-IT cluster.

e-Government Cluster, Hungary

The Hungarian government launched the Electronic Backbone (EGB) in 2004, which is a secured government-wide communication network involving 600 organizations, in order to implement its e-Government 2005 strategy. The Hungarian government organized e-Government training courses for civil servants from 700 different offices in 2006 in order to enhance internal capacities and induce a change in their mindset. It can therefore be said that the government has played an influential role in establishing IT infrastructure and capital infrastructure in Budapest, enhancing the e-Government factor conditions and inducing the creation of e-Government cluster. The public sector seems to have a more moderate influence on e-Government factor conditions in Hungary owing to the fact that very few public sector organizations actively contribute towards infrastructure development.

Business organizations were influential in setting up the e-Government cluster in Budapest. The region had numerous SMEs involved in software development, which enriched the region with skilled manpower. Also, the presence of well-reputed training institutions made the region more favorable for the setting up of the e-Government cluster. The managing agent, Albacomp Zrt., is a leader in the supply of hardware and PCs in Hungary, and could easily enter into partnerships with various other organizations in order to market e-Government solutions as Services rather than as Products. Thus, it can be deduced that private business organizations influenced in a major way to create favorable e-Government factor conditions.

GAUDI Cluster, European Union

Since the GAUDI cluster is a consolidation of three regional clusters across Europe, it was difficult to analyze how the e-Government factor conditions at European level were influenced by the factor diamond determinants. Instead, it was thought to consider a single regional cluster to conduct the factor diamond analysis. The case of the Catalonia region has been considered.

The Catalonia cluster seems to have benefited from the IT infrastructure set up by the government to implement its e-Government strategies. However, the influence of the public sector on e-Government factor conditions seems to be low. The city of Barcelona hosts Scytl, a leader in developing e-voting solutions, and other numerous consultancy organizations, internet service providers and IT based training institutions, which has resulted in creating a pool of skilled personnel in and around Barcelona. It can be deduced that business organizations have been influential in enhancing the e-Government factor conditions, especially on manpower skills, IT infrastructure and knowledge.

The Catalonia region has numerous reputed universities, which are seen as a reliable source of information and knowledge on various areas. Some of these universities have initiated programs and research initiatives related to e-Government. These universities often become training institutions for providing customized training to various types of users. The creation of CAT365, a public-private joint venture, has enabled users to secure digital signatures locally, avoiding the time and effort in applying from firms located outside Barcelona. The formation of a consortium of 800 municipalities, called Localret, is again a boon to the region as it has enabled the region to develop skills and expertise on e-Government, thereby, enhancing the skills of the people.

The summary of the Factor Diamond Analysis is shown in Table 3, which gives the estimated level of influence by factor diamond determinants on each cluster.

The above analysis suggests that in the case of the clusters studied in this paper, each of the four actors played in important role in terms of factors of production. This being said, while government and public sector do play a considerable role in providing basic factor conditions, their influence does not seem to be as high as business organizations in creating favorable e-Government factor conditions.

Table 3. Factor Diamond Analysis

Determinants	u-IT Cluster South Korea	e-Govt. Cluster Hungary	Catalonia Cluster Spain
Government	High	High	Medium
Public Sector	High	Medium	Low
Business Organisations / Associations	High	High	High
Training & Research Organisations	High	High	High

CONCLUSION

Electronic Government can be a more productive version of government in general, if it is well implemented and managed [Riley, 2003]. The concept of clusters has become widely known to induce productivity, innovation and competitiveness, which are key factors to sustaining regional economic growth and development [Porter, 1990]. Since the policies and regulations framed by government have a proven impact on cluster operations, they require being cluster-friendly and innovation-oriented. The network-based governance approach in the demand and factor diamonds of the e-Government Cluster Framework proposed in this paper, appears to be important for the success of these clusters.

Clusters often emerge and begin to grow naturally, with little role of government policies [Porter, 1998]. It is after a certain stage, that local, regional or national governments can facilitate the growth of firms within the cluster, to help them improve their competitiveness. In the case of e-Government clusters, the role of government and the importance of collaboration between various private and public partners appear much more pronounced. The internet age has seen the birth of numerous virtual organizations and virtual partnerships wherein firms collaborate and compete using online tools on the internet. Although internet facilitates the efficient exchange of information, it appears that the geographical proximity of firms cannot be avoided as this proximity facilitates the information and knowledge transfer between partners, and enhances trust and confidence within the cluster (Torre 2008).

In this paper, we have extended and adapted Porter's (1990) cluster model to the specific case of the growing e-Government sector. In particular, we hypothesized that a specific and collaborative cluster governance model needs to be in place in order to successfully create favorable demand and factor conditions for the e-Government cluster. We then examined briefly three specific e-Government cluster initiatives to give us an indication of whether various stakeholders played an important strategic role in these clusters. Our findings suggest that this was not always the case. Given the nature of the analysis, this does not, however, confirm or disconfirm the role of these stakeholders.

In conclusion, the creation of e-Government clusters is strategic in nature as it facilitates the ready availability of domain experts, the rapid selection of partners, and prompt access to innovative technologies. The pre-conditions for the creation of e-Government clusters appear to include the level of involvement of both the demand and factor diamond stakeholders as this leads to more favorable demand and factor conditions.

REFERENCES

Alberti, A., & Bertucci, G. (2006). *Replicating Innovations in Governance: An Overview. Innovations in Governance and Public Administration – Replicating what works, Department of Economic and Social Affairs*. New York: United Nations.

Altameem, T., Zairi, M., & Alshawi, S. (2006). *Critical Success Factors of E-Government: A Proposed Model for E-Government Implementation*. IEEE.

Audretsch, D. B., & Feldman, M. P. (1996). Innovative Clusters and the Industry Life Cycle. *Review of Industrial Organization*, *11*, 253–273. doi:10.1007/BF00157670

Backus, M. (2001) '*e-Governance and Developing Countries – Introduction & Examples'*, Research Report No. 3, April 2001.

Becker, J., Niehaves, B., Algermissen, L., Delfmann, P., & Falk, T. (2004). *eGovernment Success Factors*. Paper presented at the Third International Conference EGOV2004, Aug. 30-Sept. 3, 2004, Spain.

Carrincazeaux, C., Grossetti, M., & Talbot, D. (2008). Clusters, Proximities and Networks, Guest Editorial. *European Planning Studies*, *16*(5).

Chowdhury, G. H. Habib, Wahidul Md., & Kushchu, I. (2006). *Success and Failure Factors of e-Government projects implementation in developing countries: A study on the perception of government officials of Bangladesh*. Paper presented at EUROMGOV 2006.

CNN. (2008). *Can Postcards Survive the Digital Age*. Retrieved August 4, 2008 from www.cnn.com

Danish, D. (2006). The Failure of E-Government in Developing Countries: A Literature Review. *The Electronic Journal on Information Systems in Developing Countries*, *26*(7), 1–10.

Dicken, P., & Malmberg, A. (2001). Firms in Territories: A Relational Perspective. *Economic Geography*, *77*(4), 345. doi:10.2307/3594105

Gale (2006, September 27). Alien Technology Teaming with u-IT Cluster to Provide RFID Services in Korea. *Business Services Industry, Business Wire*.

Gasco, M., & Jimenez, C. (2008). e-Government and Organisational IT Adoption: The Case of the Barcelona City Council. In *European Conference on e-Government (ECEG), EPFL, Switzerland, 10-11 July 2008* (pp. 275-282).

Goldsmith, S., & Eggers, W. (2004). *Governing by Network – The New Shape of the Public Sector*. Washington, DC: Brookings Institution Press.

Heeks, R. (2001). *Understanding e-Governance for Development: A Framework for National and Donor Action*. i-Government working paper series, Institute for Development Policy and Management, University of Manchester, UK.

Heeks, R. (2001). *Understanding e-Governance for Development*. i-Government working paper series, Institute for Development Policy and Management, University of Manchester, UK.

Information for Development Program (2002, November). *e-Government Handbook for Developing Countries.* A project of InfoDev and the Centre for Democracy & Technology.

Jaeger, P. T., & Thomson, K. M. (2003). E-government around the world: Lessons, challenges, and future directions. *Government Information Quarterly*, *20*, 389–394. doi:10.1016/j.giq.2003.08.001

Korea, I. T. Times (2008, August 27). u-IT Cluster Center, a Bucket for Gold Dust. *Korea IT Times*. Retrieved from www.ittimes.co.kr

Korea e-Government (2007). *Master Plan for the Next Generation in Korea – Digital Government Inside the People*. Ministry of Government Administration and Home Affairs, Government of Korea.

Mahapatra, R., & Perumal, S. (2006, January). E-governance in India: A Strategic Framework. *International Journal for Infonomics, Special Issue 'Measuring e-Business for Development*.

Meier, M. (2001). *Finland's Forest Cluster – The Route of Competitiveness*. Tampere, Finland: Semester Paper, Tampere Polytechnic Business School.

Misuraca, G.C. (2007). *e-governance in Africa: from theory to action – A Handbook on ICTs for Local Governance*. Africa World Press, Inc.

Porter, M. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, *77*(1).

Porter, M. E. (1990). *The Competitive Advantage of Nations*. New York: The Free Press.

Riley, T. B. (2003). *E-Government vs. E-Governance: Examining the Differences in a changing Public Sector Climate*. Commonwealth Centre for E-Govenance, Canada, May 20, 2003.

Rouvinen, P., & Ylä-Anttila, P. (1999). Finnish Cluster Studies and New Industrial Policy Making. In *OECD Proceedings, Boosting Innovation: The Cluster Approach,* OECD, Paris.

Rychen, F., & Zimmermann, J. (2008). Clusters in the Global Knowledge-based Economy: Knowledge Gatekeepers and Temporary Proximity. *Regional Studies*, *42*(6), 767–776. doi:10.1080/00343400802088300

Satyanarayana, J. (2004). *e-Government – The Science of the Possible*. New Delhi: Prentice-Hall of India Pvt. Ltd.

Sund, K. J. (2006). The Geographical Concentration of Hotels in Switzerland and the Industry Life Cycle. *Tourism and Hospitality Planning & Development*, *3*(1), 1–18. doi:10.1080/14790530600638788

The Economist. (2008). *Pocket World in Figures* (2008 Edition).

The Economist Intelligence Unit. (2007). *The 2007 e-readiness rankings – Raising the bar* (White paper).

The eGov Barriers (2008). *Breaking Barriers to eGovernment – Solutions for eGovernment*. Retrieved from www.egovbarriers.org

The United Nations Secretariat. (2008). *UN E-Government Survey 2008: From E-Governance to Connected Governance*. Department of Economic and Social Affairs, Division for Public Administration and Development Management, United Nations, 2008.

Titah, R., & Barki, H. (2005). *e-Government Adoption and Acceptance: a Literature Review*. Research on Implementation and Management of Information Technology, HEC, Montreal.

Torre, A. (2008). On the Role Played by Temporary Geographical Proximity in Knowledge Transmission. *Regional Studies*, *42*(6), 869–889. doi:10.1080/00343400801922814

West, D.M. (2007, August). *Global E-Government, 2007*. Centre for Public Policy, Brown University, USA.

Zalisova, I. (2006). *Promoting Europena Regional and Local Sustainability in the Digital Economy*. European Projects and Management Agency (EPMA). Retrieved from www.atvn-eu-gp.pl/broker/resource/prelude.pdf

WEBSITES

www.albacomp.hu

www.ccegov.eu

www.ccg.gov.in

www.cluster-research.org

www.egovbarriers.org

www.egovindia.org

www.egov.vic.gov.au

www.electronicgov.net

www.epractice.eu

www.europa.eu

www.europe-innova.org

www.gartner.com

www.ifg.cc

www.nisg.org

www.politech-institute.org

www.unpan.org

www.weforum.org

www.worldbank.org

KEY TERMS AND DEFINITIONS

Cluster: A geographic concentration of inter-related businesses and institutions active in a particular industry or field.

e-Governance: The use of information and communication technology as a tool to achieving better governance.

e-Government: The application of information and communication technology solutions by the government, particularly with a view to better serving citizens.

Framework: The structure underlying a system.

GAUDI: A European cluster project launched as a consolidation of three regional clusters.

Regional Development: The provision of guidance, aid and institutional structures to support the socio-economic development of a geographical region.

U-IT Cluster: The name given to a cluster initiative sponsored by the Korean government and Incheon Metropolitan City to create infrastructure in the Songdo District (Incheon) for the development of RFID/USN technology.

This work was previously published in Global Strategy and Practice of E-Governance: Examples from Around the World, edited by Danilo Piaggesi, Kristian Sund and Walter Castelnovo, pp. 443-463, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 80
Hybrid Algorithms for Manufacturing Rescheduling:
Customised vs. Commodity Production

Luisa Huaccho Huatuco
University of Leeds, UK

Ani Calinescu
University of Oxford, UK

ABSTRACT

This chapter investigates manufacturing rescheduling of customised production and compares the results with those found for commodity production in earlier research by the authors. The hybrid rescheduling algorithms presented in this chapter were obtained by combining two key rescheduling-related elements found in the literature (a) rescheduling criteria (i.e., job priority, machine utilisation and right-shift delay) with (b) level of disruption transmitted to the shop-floor due to rescheduling (i.e., High disruption and Low disruption). The main advantage of hybrid rescheduling algorithms over individual rescheduling algorithms consists of their ability to combine the main features of two different algorithms, in order to achieve enhanced performance, depending on the objective of the organisation. The five hybrid rescheduling algorithms taken into account in this chapter are: Priority High, Priority Low, Utilisation High, Utilisation Low and Right-Shift. The authors' case study research in three manufacturing companies has identified the use of a set of these hybrid algorithms in practice. Each of the case studies is evaluated in terms of time-based performance in three main areas: suppliers' interface, internal production and customers' interface. This evaluation is carried out for both customised and commodity production, using the same hybrid rescheduling algorithms and performance measure the authors used in their previous research work, for comparability purposes (i.e. the entropic-related complexity). The findings show that customised production exhibits a lower entropic-related complexity than commodity production. Although this behaviour may seem unexpected, the entropic-related complexity analysis allows for an interpretation / understanding of its underlying reasons. For example, companies making customised products first agreed the specifications of the products with the customer, and then they mutually agreed

DOI: 10.4018/978-1-4666-1945-6.ch080

on a contract which would financially protect manufacturers (should last minute customer changes occur), by specifying analytically determined penalties or premium charges. Furthermore, a set of recommendations were made to the companies involved in this research study based on the analysis presented in this chapter, such as the need for manufacturing organisations of customised products to ensure they have dependable suppliers, and that, internally, they plan for and embed sufficient spare capacity to cope with internal or external disturbances.

INTRODUCTION

This chapter aims to assess the relationship between hybrid rescheduling algorithms, entropic-related complexity, and customised production, by using real-world manufacturing case studies. The paper thus makes a theoretical and applied contribution on these inter-related topics, which (to our knowledge) have not been previously studied in conjunction before.

The main research question explored in this chapter is: *In the context of customised production, how do hybrid rescheduling algorithms impact entropic-related complexity?* The following objectives guide this chapter: (a) To identify the typical hybrid rescheduling algorithms used in the context of customised production, and (b) To explain why and how hybrid rescheduling algorithms vary across organisations.

Given the current climate of increased global competition, manufacturing companies need to focus on customised production. It is important to consider the ever-increasing need for value-adding product design and manufacturing processes (Browning *et al*., 2002). Tu *et al.* (2001) argue that firms need to move from the internal efficiency maximisation mindset towards the emphasis on customer value. In the same vein, Professor ElMaraghy (ElMaraghy, 2009) states that one of the key challenges that manufacturing organisations face nowadays is "to satisfy the market need for products variations and customization, utilizing new technologies, while reducing the resulting variations in their manufacturing and associated cost" (p. v).

In order to satisfy the customization need of the market at a competitive price, it is necessary to understand that the above goals are neither straightforward nor easily achievable. Furthermore, the additional complexity that arises in trying to pursue them should be carefully managed. As Griffiths and Margetts (2000) point out: "customers want high quality products and services, at a reasonable cost, and they want them 'now'" (p. 155). Managing the complexity resulting from such a dynamic environment plays a key role in keeping costs under control. If organisations do not manage complexity through rescheduling or other complexity management approaches, they could face some of the following consequences (Huaccho Huatuco, 2003): customer dissatisfaction, which can then lead to losing customer demand and, related to this, less flexibility and product variety.

The type of rescheduling problem tackled in this chapter could be classified as a "stochastic scheduling problem" (Pinedo, 2008) where the disturbances were arbitrarily assigned, but the spare capacity of the original production schedule (processing times, number of jobs and number of loaded machines) varied according to a random probability distribution. These experiments were designed, run and tested in our previous work (Huaccho Hautuco *et al.*, 2009), so their detailed discussion is outside the scope of this chapter. The aim in this chapter is to provide manufacturing organisations that make customised products with recommendations on which hybrid rescheduling algorithms are more likely to be of effective use to them.

The hybrid rescheduling algorithms presented in this chapter combine priority dispatch rules with the disruption effects on the shop floor, namely: *Priority High*, *Priority Low, Utilisation High, Utilisation Low* and *Right-Shift*. These algorithms were previously studied by the authors (Huaccho Huatuco *et al.*, 2009) for commodity production. The novelty and thus contribution of this chapter consists of the extension of these hybrid algorithms to the context of customised production.

The remainder of this chapter is organised as follows. A literature review follows next, on three inter-related topics: hybrid rescheduling algorithms, entropic-related complexity and customised production. Then, the measurement methodology section gives details on the case-based research carried out. Next, the data analysis section is briefly presented. After that, the section on case studies for customised production covers the following results: entropic-related complexity, hybrid rescheduling algorithms, customised versus commodity production, and recommendations to case study companies. Next, the discussion of results section focuses on implications to Scheduling practice as well as Value-adding (VA) versus non-value adding complexity (NVA). Finally, some conclusions including some implications for theory and practice, together with some future work directions are provided.

LITERATURE REVIEW

This section discusses the links and dependencies among hybrid rescheduling algorithms, entropic-related complexity and customised production. Major, established and more recent publications in these topics include a diverse range of rescheduling methods (Ouelhadj & Petrovic, 2009; Silva *et al.*, 2008), as well as links with customised production (Da Silveira *et al*., 2001; Smart, 2009).

Hybrid Rescheduling Algorithms

There has been no shortage of hybrid algorithms reported in the literature for manufacturing systems. For example, Bierwirth & Mattfeld (1999) used Genetic Algorithms (GA) together with a tunable parameter (in terms of the length of machine idle time) to carry out hybrid scheduling, whereas Gao *et al.* (2009) also proposed the use of GA in combination with Ant-colony algorithms to determine the order of manufacturing jobs and machine assignment, respectively. Wang *et al.* (2006, 2010) focused on hybrid algorithms for the blocking flow shop scheduling problem, proposing a combination of genetic algorithms with local search algorithms. In relation to rescheduling, the work by Deblaere *et al.* (2010) combined exact reactive scheduling with Tabu search algorithms. Most of the work presented in previous literature uses simulations as a means to test the performance of rescheduling algorithms. This chapter analyses and discusses hybrid rescheduling algorithms using real-world manufacturing case studies, rather than computer simulations.

The hybrid rescheduling algorithms presented in this chapter were obtained by combining two key rescheduling-related elements found in the literature (a) rescheduling criteria (i.e., job priority) (Yamamoto & Nof, 1985; Jain and ElMaraghy, 1997, Smith 2002), machine utilisation (Jain & ElMaraghy, 1997) and right-shift delay (Yamamoto & Nof, 1985; Cheng, 1998; Holthaus, 1999) with (b) level of disruption transmitted to the shop-floor due to rescheduling (i.e., High disruption) (Yamamoto & Nof, 1985) and Low disruption (Smith, 2002). It is worth mentioning that Priority is normally assigned by order / job due date, but it could be also assigned by customer importance, by order value, or by a factor determined at the discretion of the scheduler.

The main advantage of using hybrid rescheduling algorithms is that they provide a better approach to rescheduling than if separate individual rescheduling algorithms were used, such as their ability to combine the best features of two different algorithms in order to achieve an enhanced performance, depending on the objective of the organisation. For example, while some organisations may strive on getting robust schedules (schedules that would remain valid for a long period of time), other organisations could be interested in getting near-optimal schedules which could be generated in a relatively shorter time. As mentioned before, the five hybrid rescheduling algorithms that have been used in our previous research are: *Priority High, Priority Low, Utilisation High, Utilisation Low* and *Right-Shift* (Huaccho Huatuco *et al.*, 2009).

Algorithms Description

Each of these hybrid rescheduling algorithms is briefly described next.

- *Priority High*: This algorithm takes into account the Priority of the jobs (e.g. each job's due date). Additionally, in the case of a disturbance affecting production (e.g., customer changes), the *Priority High* algorithm generates a new schedule (rescheduling) immediately. This causes high disruption to the shop floor (i.e., no threshold condition is used to filter down the disruption transmitted to the production shop floor).
- *Priority Low*: This algorithm includes a threshold condition (added to the *Priority High* algorithm) in order to filter down the disruption to the shop floor (Low disruption). In our previous research on the impact of machine breakdowns on rescheduling, the threshold condition was based on the *downtime* being greater than the *remaining* time of the disrupted job (Huaccho Huatuco, 2009). Once the threshold condition is fulfilled or surpassed (Pfeiffer *et al.*, 2007), rescheduling takes place.
- *Utilisation High*: The algorithm takes into account machine utilisation. In the case of a disturbance, it chooses the least utilized machine immediately (High disruption) to allocate the affected job.
- *Utilisation Low*: The algorithm includes a threshold condition (added to the *Utilisation High* algorithm), as explained for the *Priority Low* algorithm.
- *Right-Shift*: The algorithm delays the disrupted job until the disturbance has ended, right-shifting (delaying) all the remaining jobs accordingly.

The pseudocode for these hybrid rescheduling algorithms is provided next, where:

- **Active:** The processing time of the job before the occurrence of the disturbance.
- **Affected Jobs Downtime:** The procedure that checks that the arrival times of the jobs affected by the disturbance is right-shifted.
- **Alternatives:** The procedure that looks for alternative machines for the jobs affected by disturbances.
- **Knock-On Effect:** The procedure that right-shifts the jobs both on the affected machine and on the alternative machine, due to the occurrence of a disturbance.
- **Remaining:** The remaining processing time of the job that needs to complete after the occurrence of a disturbance.
- **Disturbances:** The procedure that writes the disturbance effect on the production schedule, in the form of unavailability of the affected machine.
- **Right-Shift:** The procedure that delays the jobs affected by disturbances.

The pseudocode for the main part of the High disruption algorithms (*Priority High* and *Utilisation High*) is given next.

The pseudocode for the Low disruption algorithms (*Priority Low* and *Utilisation Low*) is

Algorithm 1.

```
1.     Process schedule, part data and disturbance data
2.     Assign Excel data to VBA variables (This is how Arena 5.0 handles the
introduction of the analysts' own logic into the simulation package)
3.     For all jobs (i=1 to n, where n is the number of scheduled jobs)
If Job_i is affected by a disturbance, then
Rewrite arrival time of Job_i
Call 'Knock-on effect' procedure
4.     If disturbance affects a machine that is busy, then
Generate Schedule [Perform Rescheduling]
Go to (6)
5.     Process (run jobs) according to Process Schedule, then go to (10)
6.     If active_i>0, then
7.     If remaining_i>0, then
Rewrite Process Time of Part Data
Update schedule
Call 'Alternatives' procedure
Call 'Disturbances' procedure
8.     Else [remaining_i >0]
Process according to Process Schedule
Call 'Affected jobs downtime'
Call 'Disturbances'
Call 'Knock-on effect'
9.     Else [active_i >0] Call 'Alternatives' procedure
10.     End
```

given next. Note that only Step 7 changes in relation to the High disruption pseudocode.

The pseudocode for the *Right-shift* algorithm is given next. Note that Steps 7, 8 and 9 change in relation to the High disruption pseudo code.

Although disturbances can affect both customised and commodity production, the analysis of the performance needs to focus on the type of disturbance that is more relevant to the problem under study. Thus, the analysis in this chapter focuses on customer changes, which are more relevant to customised production than machine breakdowns (as studied previously). Customer satisfaction is one of the main targets of manufacturing organisations, which can be achieved by accommodating customer changes requests in their production schedule. As stated by Efstathiou (1996), schedules should retain flexibility (which could be achieved by allowing spare capacity in the production schedule, for example) in order to allow manufacturing companies attending to those requests, while at the same time companies should be able to estimate the inconvenience and the possible disruption to other orders. Customer changes include: changes in order specifications (e.g. quantity ordered [more or less], delivery times [earlier or later], cancellations, inserting new—ordinary or rush—orders) or in terms of customisation changes, modifying the product specifications. However, satisfying the customer has to be carefully balanced with production schedule stability. For example, in the case of rush orders (also called Fast Turn Around or FTA orders), Krajewski *et al.* (2005) suggested that

Algorithm 2.

```
7.     If remaining_i>0, then
Rewrite Process Time of Part Data
Update schedule
If Downtime_i>Remaining_i, then
Call 'Alternatives' procedure
Call 'Disturbances' procedure
Else [Downtime_i>Remaining_i]
Call 'Right-shift' procedure
Call 'Disturbances' procedure
```

Algorithm 3.

```
7.     If remaining_i>0, then
Rewrite Process Time of Part Data
Update schedule
Call 'Right-shift' procedure
Call 'Disturbances' procedure
8.     Else [remaining_i >0]
Process according to Process Schedule
Call 'Affected jobs downtime'
Call 'Disturbances'
Call 'Knock-on effect'
9.     End
```

suppliers can use their contracts to their advantage (e.g. charging a premium for them).

The decision element of the rescheduling model will be whether to satisfy a customer that requires a rescheduling action or not and, if so, to what extent (this is determined by the disruption threshold conditions). The threshold conditions are taken into consideration by the level of disruption to the shop floor (High or Low disruption) that the scheduler would prefer / allow to accommodate. As Monostori *et al.* (2007) expressed, finding the appropriate threshold, which ideally should be an informed and quantitative-based decision, may imply that there is a compromise between schedule stability and schedule quality. In another paper, Pfeifer *et al.* (2007) tested different levels of thresholds before rescheduling took place by evaluating the performance in terms of schedule efficiency and schedule stability. It is worth mentioning that the level of threshold can be determined either statically or dynamically (at run time). Examples of criteria which could be used for deciding which type of threshold to use include: the type of manufacturing environment, performance objectives, information availability and information processing costs.

With reference to customer changes, the following disturbances were derived from Jain & ElMaraghy (1997): (a) Increasing the priority of existing orders (as noted earlier, also known as rush orders or Fast turn around or FTA), (b) Decreasing the priority of existing orders, (c) Introducing new orders, which could be either rush orders or standard duration orders, and (d) Cancelling already placed orders. Each of these is explained in turn below. Assuming that a customer's order corresponds to a single machine job, then:

A. ***Increasing the priority of existing orders (rush orders or Fast Turn around or FTA)***: This disturbance involves finding the job corresponding to the order for which the priority is increased. If the job has not been loaded to any machine, and if the machine required to process it is free, then assign the job to start immediately on it. Otherwise, pre-empt the machine and start the job immediately, all the other jobs on that machine are right-shifted accordingly.
B. ***Decreasing the priority of existing orders:*** This disturbance involves finding the job corresponding to the order for which the priority is decreased. If the job has not been loaded to any machine, and if the machine required to process it is free, then recalculate the priority of the jobs and assign a lower priority to the job, depending on the revised due date. Otherwise, pre-empt the machine and start the next higher priority job on it, the remaining part of the job gets right-shifted accordingly.

C. ***Introducing new orders, which could be either rush orders or standard duration orders:*** If it is a rush order the highest priority is assigned to its corresponding job, then the new job is treated as in part (a) above. If it is a standard duration order, then priority is assigned accordingly, and the job is inserted into the current schedule.

D. ***Cancelling already placed orders***: This disturbance involves finding the job corresponding to the cancelled order. If the job has not been loaded to any machine and if the machine required to process it is free, then do not load it. Otherwise, pre-empt the machine and cancel the remaining completion time of the job from it. Advance any jobs that were due to be processed later on that machine to finish earlier too.

Entropic-Related Complexity

Entropic-related complexity is a measure derived from entropy (Shannon, 1949). This measure was adapted and applied to manufacturing systems in the "Complexity in the Supply Chain" project by Efstathiou & Frizelle (EPSRC Grant No. GR/M57842). This project was run jointly by the University of Oxford and the University of Cambridge, in collaboration with industrial partners including: Unilever, BAE Systems and ALPLA. Complexity, within this context, is defined as the uncertainty and the variety within a system. Hence the two main characteristics of complexity which we consider in this chapter are uncertainty and variety. Variety refers to the many parts of the manufacturing system that need to be managed, whereas uncertainty refers to the unpredictability of the behaviour of each of those parts (probability of occurrence).

Entropic-related complexity is quantified into indices that highlight high complexity areas and help focus management attention to examine their causes. The higher the index the greater the complexity displayed by the system. This can be measured in different areas or levels, for example in this chapter we focus on three areas: at the suppliers' interface, internal production, and at the customers' interface. The complexity index is evaluated based on Equation 1, which gives the fundamental form for describing entropy. Here, the entropy (H) is defined as the uncertainty and variety of the system associated with a set of n events (states), where p_i is the probability of the i^{th} event occurring. The complexity index units are given in bits per state (bps).

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i) \quad (1)$$

For example, a machine can be in one of three states: "idle," "making product" or "broken down." Each state occurs with certain frequency in a given period of time, which can then be used to obtain its probability of occurrence. Assuming the machine is "idle" for 10% of the time, "making product" for 85% of the time and "broken down" for the remaining 5% of the time, then the Complexity index is calculated as follows:

$$H = -0.1 \times \log_2 0.1 - 0.85 \times \log_2 0.85 - 0.05 \times \log_2 0.05$$

$$H = 0.75bps$$

Several practical considerations need to be taken into account when taking the entropic-related complexity measurements. First, a series of observations are made of the production processes relevant to the research study. These observations include both qualitative and quantitative behaviour, in the form of managing information flows, applying formal and informal methods, such as generating production schedules, taking / making customer phone calls, or carrying out procedures in case of unexpected events. Counting the variations and weighting each by the likelihood of occurrence gives rise to the probability of each state. The overall complexity index is calculated using Equation 1, and represents the entropic-related complexity of the system or subsystems investigated.

The level of unexpected variation (which corresponds to the uncertainty) within a system, indicates the level of "out of control." A system that behaves totally predictably can be assumed to be completely "in control." As the uncertainty and thus unpredictability of the system increases, the level of control decreases. A greater understanding of the behaviour and level of controllability of a system can be achieved through measuring the deviations from what was expected to happen. The aim of the measurements is to record these variations, and to identify their causes. This method of measuring the variations can be applied to information, to material flows, and/or to monetary values. The variations can be analysed by employing a time-, quantity- or reason-based analysis, or a combination of these.

Entropic-related complexity methods can be used for analysing and understanding how manufacturing organisations behave within their supply chain. The main idea is that the delivery performance of suppliers impacts upon the manufacturing organisation, which in turn could affect its delivery performance towards its customers. In this connection, it is hypothesised that rescheduling plays a key role in managing complexity.

Customised Production

As stated by Duray *et al.* (2000), without some degree of customer involvement in the design process, a product cannot be called customised. At the customised end of the spectrum the following customisation levels are considered: Build-to-order and Make-to-order manufacturers (Gunasekaran & Ngai, 2005). They considered that responsiveness and flexibility were both the key objectives and enablers in the customised production environment.

Four quadrants were presented in Huaccho Huatuco (2003), which resulted from the combination of the X-axes: "functional product" and "innovative product" and the Y-axes: "efficient scheduling objective" and "responsive scheduling objective." The two quadrants of interest to the work presented in this chapter are: "functional product, responsive scheduling" and "innovative product, responsive scheduling." The former refers to commodity production, whereas the latter refers to customised production. In terms of managing complexity, companies belonging to either of these two quadrants were identified to manage complexity using "rescheduling." However, in the search for an alternative way of managing complexity, the emphasis shifts from using "decision-making," for commodity production, towards the use of "spare capacity" in customised production. So, ideally, the production schedules in customised production should embed higher spare capacity levels than the production schedules in commodity production. "Responsive scheduling" is related to the Just-in-Time (JIT) Operations Management philosophy by its aim to satisfy the customer by delivering the right product, in the right quantity and at the right time.

Duray *et al.* (2000) built upon Mintzberg's work (1988), and stated that customization can be pure (from the conception stage), tailored (at the fabrication stage) or standardised (at the assembly stage).

Summary of the Literature Review

This section has covered the three inter-related topics of interest for this chapter: hybrid rescheduling algorithms, entropic-related complexity and customised production. Hybrid rescheduling algorithms have been extensively used in the manufacturing context, since they are robust in providing solutions that could take the best features of each algorithm. The rescheduling algorithms considered in this chapter are: *Priority High, Priority Low, Utilisation High, Utilisation Low* and *Right-shift*. Entropic-related complexity has been presented as a holistic, comparable and useful measure to managers. They can use it to prioritise and direct their efforts in an informed and objective manner. Customised production was discussed in the light of the level of customisation, and the need to enable organisations to achieve flexibility and responsiveness. This can be done through a combination of "rescheduling" and "spare capacity" approaches to managing complexity.

MEASUREMENT METHODOLOGY

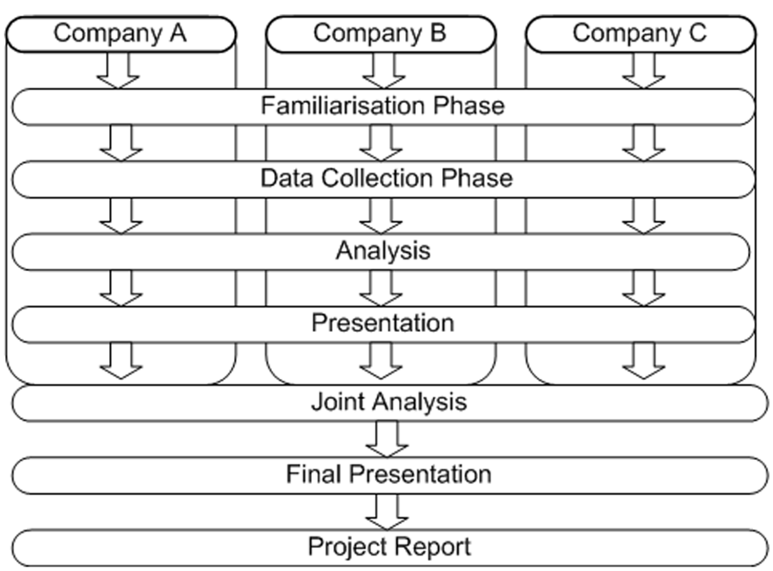
The measurement methodology used in this research was case-based research, with individual-companies case study analyses followed by a cross-case analysis. The case study protocol was followed (Voss *et al.*, 2002) with particular application of the stages proposed in Sivadasan *et al.* (2002), which are diagrammatically represented in Figure 1.

Company A, Company B and Company C participated in the "Rescheduling and complexity for customised products" project. The information flow complexity within these companies was evaluated by measuring the entropic-related complexity (i.e., the uncertainty and variety of the information transfers between suppliers and customers), for example. Data was triangulated using: (a) semi-structured interviews, (b) observations of the shop floor and (c) documents in electronic or hard copy form.

Familiarisation

Prior to conducting the actual complexity measurements, between two and three days were spent within each company for collecting preliminary information with people, information flows and processes. This involved an interactive process of semi-structured interviews with key personnel, to understand the operations and to map out the key material and information flows. In terms of plant familiarisation a tour of the plant sites was given by the companies, which provided a better understanding of the manufacturing processes such as process layout, and additional information about types of resources and their capabilities, at each organisation.

Figure 1. The case study methodology



Key Information Flows and Variations

The diagrammatic representation of the information flows studied is given in Figure 2, where the arrows represent the information flows.

The variations identified and investigated in this research are given in Table 1.

The Parameters Considered Prior to Data Collection

To ensure the effectiveness and accuracy of data collection, and the applicability of the complexity measurement method, the following parameters were determined prior to taking measurements:

- The variables to be measured (time-based, quantity-based or reason-based variables).
- The possible states of interest (using Equation 1) (e.g. at any given time a machine can be in one of three states: "idle," "making product" or "broken down").
- Frequency and duration of observations (e.g. once every hour, shift or day, for two weeks).
- Key information flows and variations (e.g. despatch notes and delivery requirements for the last three months).

The Variables to be Measured

For example, in the case of deliveries at the supplier interface, goods may not arrive on time and / or in full (OTIF) (i.e., time and quantity variations). The variables we chose were: time-based deliveries—for the suppliers' and customers' interfaces, and time-based production—for the internal production.

Additionally, the reasons for time and quantity variations were identified and recorded, in order to detect the instances of importing, exporting and absorbing complexity. Briefly, importing complexity consists of accepting complexity from elsewhere outside the organisation (e.g. suppliers or customers), whereas exporting complexity consists of transferring complexity to those organisations.

Figure 2. Information flows investigated

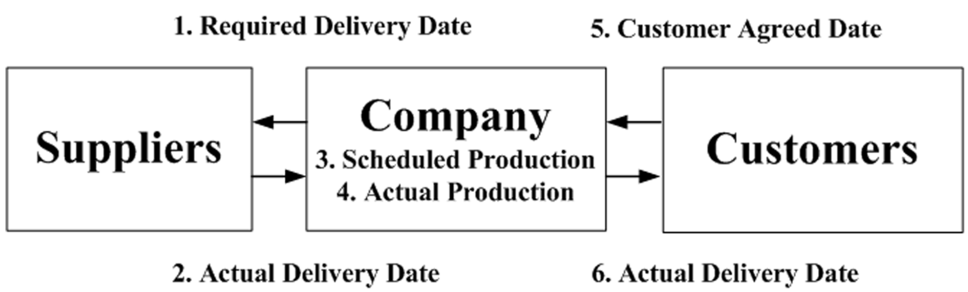


Table 1. Variations identified and measured

Area	Variation*
Suppliers' interface: Procurement	2. Actual Delivery Date – 1. Required Delivery Date
Company: Internal Production	4. Actual Production – 3. Scheduled Production
Customers' interface: Delivery	6. Actual delivery Date – 5. Customer agreed date

*Note: the symbol "–" refers to arithmetic difference.

Absorbing complexity is the way organisations decrease / minimise complexity internally. For a full discussion on complexity transfers please refer to Sivadasan *et al.* (2004). Also, the reasons for variations are useful in order to facilitate the understanding, interpretation and meaningfulness of the entropic-related complexity.

The Possible States of Interest

The results of this study focused on time-based analysis (i.e. the variations stated in Table 1). States were defined so as to accurately capture the investigated processes. They were kept the same across all case studies, whenever possible, for comparability purposes. The following states of interest were taken into account at the suppliers' and customers' interface:

- on time or early.
- < 1 week late.
- 1-2 weeks late.
- 2-4 weeks late.
- 4-8 weeks late.
- > 8 weeks late.

Frequency and Duration of Observations

The frequency of the information and material transfer was used as a guideline for the frequency of measurements. In the case of event-based activities (such as deliveries) the analyses utilise all the available data for each case study, which can be either live or historical. In terms of live data, two weeks were used for carrying out the data collection on site. In terms of historical data, it was determined that at least three months of weekly data would provide sufficient data points for the analysis.

Data Collection

Data collection consists of sampling information and material flows over time. In this study, information variations were investigated and recorded in terms of actual versus expected performance over the data collection period. Thus sufficient data (typically a minimum of 30 data points) should be gathered to ensure each state have been observed to give an estimate of its probability. As mentioned earlier, data consists of: (a) semi-structured interviews, (b) observations of the shop floor and (c) documents in electronic or hard copy form.

Data Analysis

The data analysis involved the calculation of the complexity indices, for the variations mentioned earlier, and the interpretation of the results thus obtained.

The computation of the complexity indices comprises five steps:

1. Calculate the arithmetic difference between the actual information flows and the scheduled information flows (e.g. if a delivery was expected on the 5th August, but it arrived on the 4th August, then the arithmetic difference would be: 4-5=-1 [a day early]).
2. Classify the difference calculated at step 1 using the pre-defined states (e.g. if a state is defined as "a day early") then, for illustration purposes, suppose 40 deliveries belong to that state.
3. Estimate the probability associated with each state. For the example above, if the total deliveries were 400, therefore 40 deliveries represent 10%, so the probability of "a day early" deliveries is 0.10.
4. With the estimated probabilities at step 3, use Equation 1 to calculate the entropy (e.g. continuing with the calculation from previous step: $-0.1 \log_2 0.1 = 0.33$ bps).

5. Add up all the entropy values to calculate the entropic-related complexity (e.g. 0.33+…=2.58 bps).

The subsequent analysis of the indices involves three more steps:

1. Compare relevant complexity indices calculated for the different areas being analysed, to identify high complexity areas. The rule of thumb used here was: < 1 bps: low; 1 to 2 bps: medium; > 2 bps: high. This rule of thumb results from considering that the maximum entropic complexity for a six-state system is 2.58 bps or bits per state (=$\log_2 6$), and a rough estimation of the low, medium and high values, respectively. It is worth mentioning that the specific state definitions have a direct impact on the results. However, the recommendations derived from the calculations in this chapter are robust, as they hold when tested with slightly different states, as in the case of Company B (Huaccho Huatuco *et al.*, 2010). For more details on the issue of state definitions for the entropic-related complexity, please refer to Sivadasan *et al.* (2001).
2. Identify specific reasons within each complexity area (e.g. suppliers' interface, internal production and customers' interface), and observe their frequency of recurrence throughout the flows, grouping them if necessary. For example, similar reasons can be aggregated into a state, "low materials quality" and "delays in suppliers' delivery" could be grouped into "suppliers-related reasons."
3. Provide recommendations to the manufacturing organisations about managing their entropic-related complexity. The key findings, such as high complexity areas and prioritisation of the areas that require managerial attention, were presented to each participating manufacturing organisation.

CASE STUDIES FOR CUSTOMISED PRODUCTION

This section is based on the "Rescheduling and Complexity for Customised Products" project involving three UK manufacturing companies in customised production (Huaccho Huatuco, 2006). The case studies are summarised in Table 2.

Both material flows (such as the raw materials and work in progress) and information flows (such as the production schedules) are important when rescheduling manufacturing systems. Details of the companies rescheduling practices were collected from interviews, direct observations of the shop floor and paper / electronic documents. These case studies are used to abstract the characteristics of the hybrid rescheduling algorithms as used in practice.

The following benefits to the participating companies were anticipated:

- Diagnosis and benchmarking of each company's rescheduling practice.
- Evaluation of the amount of entropic-related complexity they are handling when rescheduling.
- Recommendations as how to manage complexity to their advantage.

These companies are based in the UK and were chosen among different candidate companies due to their different levels of customisation. All three companies agreed to participate in the research project by providing access to their management staff in relation to scheduling and production.

The following good practice elements were gathered from all the case studies:

- People involvement and commitment.
- Frequent communications both formal (printed documents) and informal (phone and face-to-face conversations).
- Expertise coming from experience.
- Quality awareness throughout the process.

Table 2. Summary of the case study companies

Company (size)	**Product**	**Main disturbance(s)**	**Relative level of customisation according to Duray *et al.* (2000)**
A (Large, part of a multinational corporation)	Pumps for oil, water or waste	Suppliers do not always deliver on time. Bottlenecks exist in the production shopfloor (as a consequence, queues develop)	Each pump is a project on its own, depending on where the pump is to be installed. The customer determines the specification of the pump. Customisation is tailored (at the fabrication stage).
B (Large, part of a multinational corporation)	Paper chemicals manufacturer	Customers change their orders at the last minute, and innovative products are likely to fail in the shop floor. For example, new chemical products for paper to achieve particular properties, such as: allowing the quick drying of the ink printed on them.	Innovative paper chemicals with improved properties. Customer needs are gathered by the company's own dedicated customer service staff who work closely with customers. Customisation is pure (at the conception stage).
C (SME)	Industrial doors	None apparent	Each door is a project on its own, depending on where the door is to be installed. There is a limited range of door types on offer (four types), but the overall specifications such as size and location characteristics are unique for each door. Customisation is standardised (at the assembly stage).

Some of the issues identified at each manufacturing organisation are given in Table 3.

Table 4 shows the data collection details in terms of time period. It also shows that between 3 and 17 months historical data were collected.

The remainder of this results section is divided into: entropic-related complexity, hybrid rescheduling algorithms, customised versus commodity production, and recommendations to case study companies. These are discussed in turn next.

Entropic-Related Complexity Results

The results were shown to each manufacturing organisation by means of a Power Point presentation. Prior to this meeting, hard-copies of the presentation were distributed to the attendees to allow them to assess the confidentiality and sensitivity of the information. The attendees included key people within the organisation, people who had participated in the case study and people to whom the results would be most relevant. The participants provided feedback on their views and agreed with the results of our analysis. The most important point they made was in regards to being able to "quantify" the amount of complexity they were handling. Also, they acknowledged the fact that internal complexity does not exist in isolation and it is influenced by their suppliers' and customers' interfaces.

The results of the analysis have been classified, according to the type of entropic-related complexity handled due to rescheduling: Suppliers' interface, Internal Production, and Customers' interface. Each of these categories is discussed in turn next.

Suppliers' Interface

The complexity indices associated with the procurement or suppliers' performance are plotted in Figure 3. An example of the entropic-related complexity calculation is presented next. Company C's suppliers' interface was observed to be

Table 3. Issues identified

Company	Issues identified
Company A	• Difficult for the researchers to see the big picture (i.e. what is happening in other departments and the organisation as a whole). • Not much historical data recording. • Pressure on the scheduler for the individual projects to meet milestones / deadlines.
Company B	• Frequent rescheduling due to customer rush orders • No charge to customers for rush orders • Flexibility: strength or weakness? There is enough spare capacity to accommodate rush orders. However, the scheduler is finding increasingly difficult to manage the complexity associated with them. • Not enough operators – in the short term this is mostly because in the original schedule not many operators are needed, but in practice with the new rush orders coming in, shortage of labour becomes an issue.
Company C	• No extra charge to customers for rush orders or fast turnaround (FTA) deliveries. • Production earliness could become an issue when storage space is limited.

Table 4. Data collection details

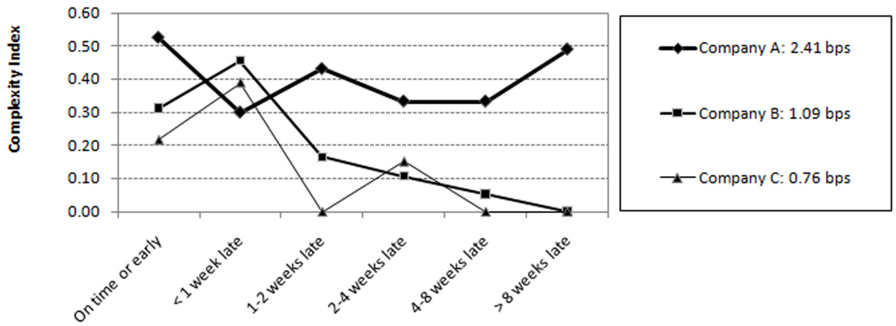
Company	Area of assessment	Time Period
Company A	Suppliers' interface	10 months
	Internal Production	3 months
	Customers' interface	17 months
Company B	Suppliers' interface	3 months
	Internal Production	9 months
	Customers' interface	3 months
Company C	Suppliers' interface	7 months
	Internal Production	4 months
	Customers' interface	4 months

in one of the following states: "on time or early," "< 1 week late" and "2-4 weeks late." There were: 21, 59 and 20 occurrences of each of these states, respectively. Applying Equation 1, they generate: 0.22, 0.39 and 0.15 bps, respectively. Adding these up, the complexity associated with the suppliers' interface at Company C is calculated at 0.76 bps.

In Figure 3, the horizontal axis shows the states of earliness or lateness associated with suppliers' deliveries to Company A, Company B and Company C, whereas the vertical axis shows the complexity index associated with those states. It can be seen that the average complexity associated with deliveries from suppliers to Company A is high at 2.41 bps (>2 bps), it is medium from suppliers to Company B at 1.09 bps (1-2 bps) and it is low from suppliers to Company C at 0.76 bps (<1 bps). All three results are within the limits of manageable complexity (Miller, 1956).

The complexity associated with the state of being "on time or early" could be regarded as value-adding complexity, which is the complexity that is needed to satisfy the end customer. However, early deliveries can cause some problems to companies, in terms of inventory, especially those working in a lean / JIT environment. The states of lateness are regarded as non-value adding complexity. Suppliers to Company A are most complex in the "on time or early" and the ">8 weeks late" states. This could be explained as due to the fact that suppliers' deliveries of common components are not a problem, whereas suppliers' deliveries of critical components such as pump motors are more prone to take longer than estimated. Suppliers to Company B and to Company C show a similar pattern of behaviour, with the highest source of complexity coming from the "<1 week late," and decreasing for the rest of the lateness states. This is because their suppliers are reasonably reliable and, if they missed the requested delivery date, they will deliver sooner rather than later. This is a direct consequence of the percentage of occurrence of states, which is presented in Table 5. As mentioned earlier, states are pre-defined in the analysis, so another way of analysing the data could consider two states: "on

Figure 3. Entropic-related complexity at the suppliers' interface



time or early" versus "late" deliveries. However, the six states used here have been defined taking into account the scheduler's decision making points as to whether to take action according to their implications on the shop floor production. For example, lateness of less than a week (i.e. "<1 week late") would not normally prompt as much managerial attention as lateness by a month or so (i.e. "4-8 weeks") would.

Internal Complexity

Although all three companies make customised products, they have their own individual production scheduling characteristics that make them different from one another. For example, their production lead time is 9 to 12 months, 1 to 2 weeks, and 4 weeks for Company A, Company B and Company C, respectively. Thus, the complexity results of their internal production are presented separately, in the next three sections.

Table 5. Percentage of occurrence of states (suppliers' interface)

State	Company A	Company B	Company C
On time or early	32	75	84
< 1 week late	8	19	13
1-2 weeks late	17	3	0
2-4 weeks late	10	2	3
4-8 weeks late	10	1	0
> 8 weeks late	23	0	0

Company A

Table 6 shows the percentage of occurrences for each of the areas which can be rescheduled in the order book. It can be observed that the schedule adherence ("OK" state) accounts for 17% of the total. The following processes contribute almost equally: factored, project equipment and materials (between 14% and 15%). Most importantly, it can be observed that rescheduling of the order book due to the assembly area accounts for 22% of the total. This is not surprising, given the nature of the product which is engineer-to-order pumps, and which implies that there is a high level of uncertainty when making one-off products.

Table 6. Order book rescheduling by production process at Company A

Production process	%
OK (production according to schedule)	17
Assembly	22
Factored	15
Project equipment	15
Materials	14
Machining	10
Fabrications	7

Figure 4 shows that the amount of complexity that is handled at Company A accounts for 2.80 bps, which is classified as high in this research. The value-adding complexity (represented by the black bar) accounts for 15% of the total, and the state that contributes the most to the non-value adding complexity is "assembly." Assembly consists of a series of steps, some of which are outsourced, so the performance is highly influenced by the performance of the outsourcees.

Company B

Table 7 shows that on time or according to the schedule production accounts for 46% of the total. The data used in the table corresponds to the combined live and historical data. Also, it can be seen that the percentages increase with lateness (i.e. once Company B has missed the production scheduled date, the customised product is expected to be made later rather than sooner). This can be due to technical issues, such as the different behaviour of the chemicals in the shop floor compared with the experiments carried out previously in the laboratory.

In Figure 5, we observe that Company B manages a high overall amount of complexity, at 2.16 bps, and that the value-adding complexity (represented by the black bar) accounts for 24% of the total. Then the non-value adding complexity states increase with lateness, with the biggest one coming from the ">8 days late" state. This can be attributed to technical errors during production and to knock-on effects in the production schedule.

Table 7. Actual versus scheduled production performance at Company B

State	%
On time	46
<2 days late	6
2-3 days late	6
3-6 days late	12
6-8 days late	12
>8 days late	18

Company C

During the period for which data were collected, which comprised both live and historical data, Company C's production was always on time or early. This was mainly due to the fact that the standard delivery time had been set to four weeks, which gave plenty of time to account for any unforeseen circumstances that could delay production. So, Company C's way of absorbing complexity consisted of allocating high spare capacity levels in their original production schedule. This also relates to the "spare capacity" identified earlier as another way of managing complexity

Figure 4. Internal entropic-related complexity at Company A (2.80bps)

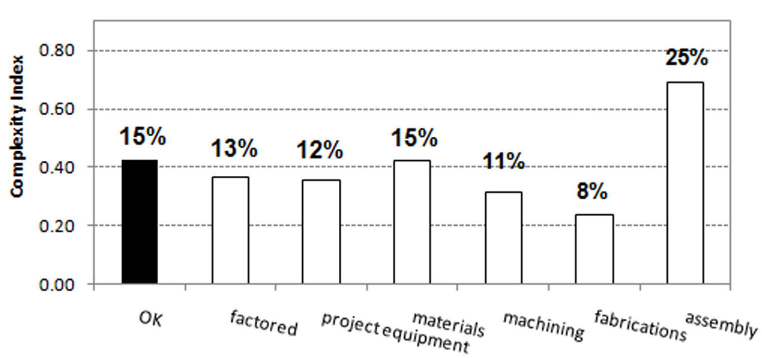
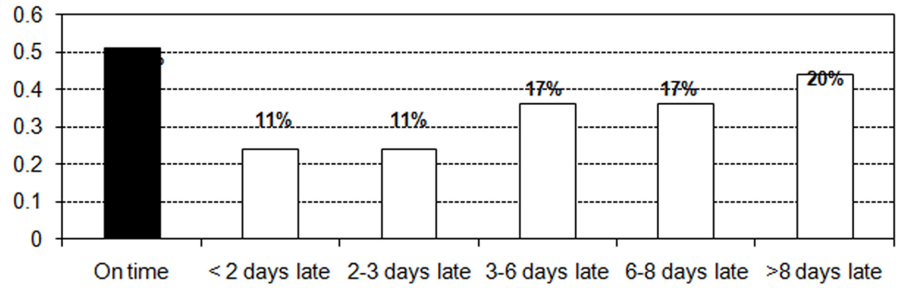


Figure 5. Internal entropic-related complexity at Company B (2.16bps)



instead of "rescheduling." It was also noted that Company C placed a strong emphasis on fixed delivery dates with the customer, so once these are agreed with them, deadlines are quite firm, with the customer potentially bearing the costs for any rescheduling caused by them.

Further analysis was performed regarding the earliness of production, as seen in Table 8. It can be observed from this table that 50% of the doors were finished up to a week early; another observation is that the percentage of finished doors decreased with increased levels of earliness.

The overall amount of complexity handled at Company C, at 1.99 bps is in the medium range (1-2 bps), as seen in Figure 6. The value-adding complexity or "on-time" delivery accounts for 21% of the total. It could be argued that finishing products early is value-adding, since once the doors are completed the resulting spare capacity could be used for accommodating other orders. However, finishing products too early, such as the last three states ("9-11," "11-13" and ">13" days early) may be tying up capital that could be otherwise used, doors could become damaged or lost, and physical space may become an issue if the business were to expand. That is why those early states have been classified as non-value adding. This understanding was communicated to and validated by the company during the presentation of results.

Table 8. Actual versus scheduled production performance at Company C

State	%
On time	16
<7 days early	50
7-9 days early	19
9-11 days early	5
11-13 days early	5
>13 days early	5

It is worth mentioning that Company C's performance is remarkable, because no instances of late internal production were identified during the analysis period. This case is unusual among the other case studies carried out to date by the authors. The fact that 74% of Company's C requested deliveries were classified as fast turnaround (FTA) or rush orders makes its reliable performance even more surprising.

Customers' Interface

The entropic-related complexity was calculated at the customers' interface too. Table 9 shows the percentage of occurrences for each state at each company. It can be seen in this table that Company A and Company B exhibit about 60% on time or early deliveries to customers, compared with 98% from Company C. It is also shown that the last two lateness states ("4-8" and ">8 weeks" late) are heavily occurring in the case of Company A, but are hardly seen in Company B or Company C.

Figure 6. Internal entropic-related complexity at Company C (1.99bps)

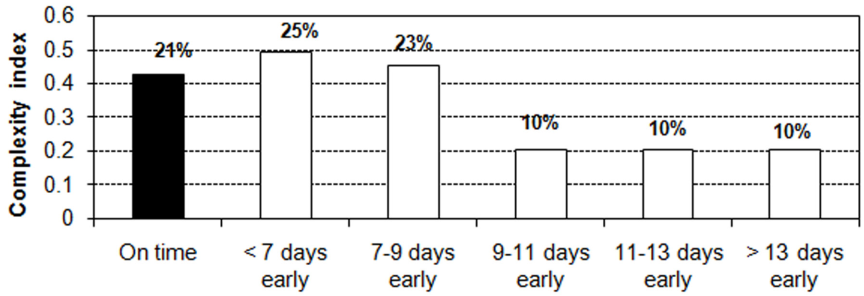


Table 9. Percentage of occurrence of states (customers' interface)

State	Company A	Company B	Company C
On time or early	61	62	98
< 1 week late	2	20	2
1-2 weeks late	5	9	0
2-4 weeks late	7	7	0
4-8 weeks late	10	1	0
> 8 weeks late	14	1	0

In this context, Figure 7 shows that Company A and Company B exhibit a medium amount of complexity (1-2 bps), whereas Company C transfers a very low amount (<1 bps). Company B and Company C show a similar pattern in terms of the states, with the highest being "<1 week late," and then decreasing complexity as the lateness increase. This means that once these companies have missed a promised delivery date they will try to deliver sooner rather than later. Company A's deliveries show a different pattern, with a distinctive increase of the complexity as lateness increases (Figure 7). The reason for this behavioural pattern is that, having missed the deadline, efforts are made to deliver as soon as possible, to avoid reaching and surpassing the penalties ceiling. Once this ceiling is surpassed, the efforts are then directed towards other projects (pumps) that may be in danger of becoming late, too.

Summary of Entropic-Related Complexity Results

Table 10 summarises the results presented so far in this chapter. The results indicate that the higher the complexity at the suppliers' interface, the higher the complexity handled internally. As mentioned earlier, the more complexity is imported by suppliers or customers into the organisation, the more complexity the company would need to handle internally. So, not surprisingly, good suppliers enable good internal performance. It was also noticed that all three companies handled less complexity at their customers' interface than the complexity they handled internally. This can be explained as the case study companies absorbing complexity (i.e., using ways of managing complexity such as "rescheduling" and the use of "spare capacity" in the production schedule to absorb it). It was suggested to companies that they should make their customers aware of this fact. By doing this, customers could become more amenable to pay a Premium (cost of peace of mind), as in the case of Company C; this approach could also help ease some of the pressure from Company A. These understanding and control mechanisms can be linked to the variety and uncertainty of processes, which in the case of customised products were prone to happen more often, as seen in Company B.

Figure 7. Entropic-related complexity at the customers' interface

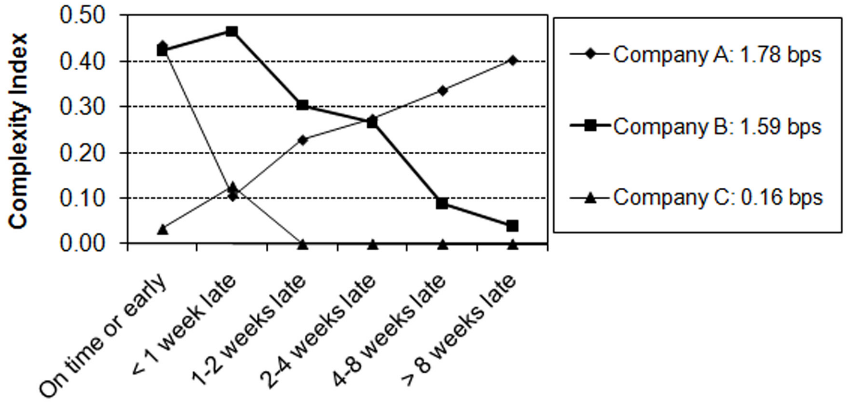


Table 10. Entropic-related complexity in bits per state (bps)

Case study	Suppliers' interface	Internal complexity (value adding %)	Customers' interface
Company A	2.41	2.80 (15%)	1.78
Company B	1.09	2.16 (24%)	1.59
Company C	0.76	1.99 (21%)	0.16

Note: <1:low; 1-2: medium; >2: high

Hybrid Rescheduling Algorithms Results

Company A

The emphasis of the production of the pumps at Company A was on the deadlines for achieving the milestones of each project (pump). If a need for rescheduling occurred, Priority was given according to the importance of the project (dictated by its monetary value). Company A operates in a highly regulated supply chain environment, with contractual clauses which will penalise it for not delivering on time to the customer. In terms of the level of disruption transmitted to the shop floor, the scheduler at Company A would consider the customer requests carefully against all the projects that were in the pipeline, and would just delay the jobs rather than pre-empt and interrupt the course of the jobs already running. So, Low disruption was transmitted to the shopfloor. Therefore, it could be concluded that Company A's preferred hybrid rescheduling algorithm was Priority Low (Figure 4).

Company A's poor performance at the customer interface is mostly the result of a knock-on accumulated effect of its suppliers' failure to deliver on time and in full, and of its own NVA internal complexity.

Company B

In terms of rescheduling strategies, it was observed that Company B was using Utilisation-based rescheduling strategies. This is because it was normally operating with large amounts of spare capacity, in the form of idle vessels. Thus, when a rush order was requested by a customer, Company B could easily accommodate it in its current production schedule by allocating the rush order to one of the vessels that was available. In terms of disruptions transmitted to the shopfloor, the objective of the scheduler was to satisfy the customer. As a result, the production schedule was updated quite often, on a shift-by-shift basis. So, it could be concluded that Company B's preferred hybrid rescheduling algorithm was Utilisation High (Please note that "High" refers to the high disruption transmitted to the shopfloor and not to high utilisation). See Figure 5.

In terms of the disturbances due to customer requests, Company B showed the following common requests: (a) increasing the priority of existing orders, and (b) introducing new orders, which were mostly rush orders. From the interviews at Company B it was found that it had verbal agreements with suppliers and contractual agreements with customers. This was because, as Company B's customers are a small set of paper mills (who may switch to another supplier if not satisfied), they had to pay extra attention to customer service. This could place Company B in a vulnerable situation in the context of its supply chain, since power can be imbalanced (New & Ramsay, 1997).

Company C

Company C did not need to carry out rescheduling at all (Figure 6) – the reasons for this are explained next. Firstly, at the suppliers' interface, Company C carefully selected its suppliers, such that they were highly dependable. It is worth mentioning that Company C was in a position to do so, as they focussed on a niche in the market, raising issues of costs versus added-value. Internally, spare capacity was built in the production schedule, so it remained mostly unaffected by disturbances (again a way of absorbing complexity). At the customers' interface, Company C used formal contracts with customers to protect itself against last minute changes. As a consequence, Company C did not need to reschedule its production at all during the studied period. This result was due to including spare capacity in its production schedule, and to its careful selection of suppliers and customers. This solution may be more suitable and feasible in the Company's C industry type rather than in other industries.

Customised vs. Commodity Production

Table 11 deals with the comparison between customised and commodity production results. Key differences between these two types of production on observed rescheduling practices are:

- The use of threshold conditions (associated with Low disruption rescheduling algorithms) for carrying out rescheduling is more common in customised production than in commodity production. Threshold conditions mean the use of decision making points in order to filter disruption transmitted to the shop floor due to disturbances (Low disruption). Typically companies of

Table 11. Rescheduling characteristics: customised vs. commodity production

Rescheduling characteristics	Customised production	Commodity production
Product unit cost (£)	High	Low
Product type	Specialised	Standard
Pace of production	Slow	Fast
Threshold conditions used for production rescheduling	More common	Less common
Spare capacity embedded in the production schedule	Medium to High	Low to Medium
Company protected by contractual agreements with customers	Very common	Not common
Preferred hybrid rescheduling algorithms for reducing entropic-related complexity	*Priority Low* (Company A) and *Utilisation High* (Company B)	*Utilisation High* and *Right-shift* (Huaccho Huatuco *et al.*, 2009)

customised production are protected by formal contracts against customer changes, with attached penalties / premiums for last minute changes or late deliveries.
- Spare capacity in the original production schedule is higher in the customised production than in commodity production.

A key similarity between the two contexts is that *Utilisation High* is a useful hybrid rescheduling algorithm in both customised and commodity production. Additionally, *Priority Low* (as observed in Company A) is better suited for customised production, and *Right-shift* is better suited for commodity production.

RECOMMENDATIONS TO CASE STUDY COMPANIES

The following recommendations were given to the case study companies; these are linked with Table 3, about the issues identified:

Suppliers' Interface (Figure 3)

- Company A (high complexity): Exercise tighter control over suppliers' performance, develop partnerships with key reliable suppliers, and allow the production schedulers to play a more important role in the process of purchasing materials or components. This approach will allow Company A to focus on managing its own internal complexity, by first reducing the uncertainty of suppliers' deliveries. However, it should be kept in mind that changing structures / responsibilities could be difficult in a large company such as this. Change should be driven by the top management with direct involvement of the shopfloor management and operators, and it should be carefully monitored and controlled to ensure that the overall objectives are achieved.
- Company B (medium complexity) and Company C (low complexity): Maintain the good planning of purchasing materials by the production scheduler, and further develop good relationships with reliable suppliers. These were identified through observations / interviews with the productions schedulers at these companies.

Internal Complexity (Figures 4, 5 and 6)

- Company A (high complexity): Synchronise production machining-assembly-test, in order to avoid bottlenecks, and improve forecasting of delivery dates (they were too optimistic) by allowing some slack in the production schedule, to cater for unforeseen problems (this recommendation also applies to Company B). This may allow the schedule to become more robust, therefore reducing the need for rescheduling. However, it is acknowledged that they may find allocating slack time in the schedule is too costly, if there are idle times when many jobs are waiting to be processed.
- Company B (high complexity): Continue monitoring the production system; rescheduling does help towards delivering the products on time.
- Company C (medium complexity): Improve forecast of delivery dates (they were too pessimistic), since most of the complexity is coming from earliness. This is a potential problem, and the company was made aware of this at the result presentation. This may seem to contradict the earlier "slack in the schedule" recommendation. However, it was noticed that Company C was overbuffering itself, and it could easily promise the customer and achieve "on time" deliveries of standard

products up to 3 weeks instead of 4 weeks, should they wish to do so.

Customers' Interface (Figure 7)

- Company A (medium complexity): Develop partnerships with customers, with especial review of the contractual agreements and the applied penalties in case of lateness.
- Company B (medium complexity) and Company C (low complexity): Develop threshold conditions regarding rush orders or fast turnaround products (FTA). This will allow Company B to achieve Low disruption transmitted to the shop floor because, during the period of study, the customers were taking for granted that their requests will be taken into account without penalty. For Company C, the threshold is so wide that it could risk becoming uncompetitive, should its competitors offer to provide the same product faster.

Although managers knew that certain areas were more complex than others, entropic-related complexity allowed them to measure and to compare the complexity levels. The ability to integrate and quantify complexity in a single comparable measure was considered useful by managers, and it provided a basis for fruitful discussions between researchers and practitioners.

DISCUSSION OF RESULTS

The counter-intuitive result in this chapter is that companies in customised production do not necessarily face high levels of complexity. This is because, as observed in the companies reported in this chapter, in practice companies in customised production ensure that they have formal contractual agreements in place for the delivery of the final products. Also, these companies ensure they have the resources to enable them flexibility, and to cope with the additional process complexity due to customisation. This can work in their favour, such in the case of Company C, which uses spare capacity as a buffer to protect itself from disturbances. Alternatively, it could work against the company, such as in the case of Company A, which tries to avoid the lateness penalties of its tightly arranged schedule. It could be argued that Company C operated in a niche market and could afford to place higher conditions onto its customers, than, for example, Company B, for which competition is more intense.

Table 12 compares the results previously reported in the literature by the authors (Huaccho Huatuco et *al.*, 2009) with the results of the case studies reported in this chapter. It can be seen that *Utilisation High* achieves lower levels of entropic-related complexity than *Priority Low*, in both customised and commodity productions. However, customised production shows, surprisingly, a lower level of complexity than commodity production for both identified hybrid rescheduling algorithms. In order to explain this, it is important to use the case-based research methodology to investigate entropic-related complexity, since it also captures the qualitative aspects through interviews and observations. Direct observations and historical data have been used to obtain meaningful results, by providing explanations from schedulers and managers about the reasons for rescheduling. Historical data can also help to smooth out any unusual performance into a more stable, and thus representative pattern.

Implications for Scheduling Practice

The more threshold conditions are embedded in the hybrid rescheduling algorithm, the less disruption is transmitted to the shop floor. So, this chapter encourages manufacturing organisations to be reactive to disturbances, whilst also taking into account thresholds for deciding when to re-

Table 12. Entropic-related complexity for each hybrid rescheduling algorithm: customised vs. commodity production

Hybrid rescheduling algorithm	Customised production	Commodity production*
Priority High	Not identified	2.731[b]
Priority Low	2.800	3.028[a]
Utilisation High	2.160	2.709 [b]
Utilisation Low	Not identified	2.713 [b]
Right Shift	Not identified	2.712 [b]

*Taken from Huaccho Huatuco *et al.* (2009): (a) The complexity associated with the information content of schedules and (b) The complexity associated with the variation between schedules (bps).

act. As discussed in the literature review section (Monostori *et al*., 2007; Pfeiffer *et al*., 2007), thresholds could be tested at different levels using computer simulations in order to test 'what if' scenarios and taking into account the studied organisation's objectives.

The High disruption algorithms (*Priority High* and *Utilisation High*) monitor the state of the system and react whenever there is a disruption. This is called event-based rescheduling. However, that reaction is constrained sometimes because the alternative machine is busy, but the effort to check alternative machine availability had already been made. Whereas the Low disruption algorithms (*Priority Low* and *Utilisation Low*) monitor the state of the system, they only react if the remaining processing time is greater than a threshold. When executing the *Right-shift* algorithm there is less need to update the schedule, as it only affects the jobs on the affected machine.

Value-Adding (VA) vs. Non-Value Adding Complexity (NVA)

Table 10 depicts the value adding percentage of internal complexity. In this chapter value adding is related to "on time or early" production, as it is assumed that these states will allow a manufacturing company to satisfy its customers. In this respect, Company B and C achieved over 20% VA complexity, whereas Company A achieved a 15% VA complexity.

It is worth mentioning that determining what is VA and NVA complexity depends on the company or industry sector under study. A trade-off of the potential benefits versus losses should be performed in order to give more general insights and to decide the VA and NVA states and complexity.

Even when a company is handling a high amount of VA complexity, this still needs to be managed by the company. However, a different type of action is needed for this. As mentioned in the Appendix: "Key Terms and Definitions," the manager would need to decide on the VA complexity level it could allow into the system depending on the company, sector and wider context.

The NVA complexity, which is the complexity that does not help with satisfying the customers, needs to be both monitored and controlled. NVA complexity could possibly not be entirely eliminated, but it can be reduced. In order to do this, the reasons that cause it in the first place need to be identified and analysed.

CONCLUSION

The conclusions section provides theoretical-based insights for customised versus commodity production, and it derives practical guidelines for production managers / schedulers about the hybrid rescheduling algorithms which are more likely to succeed in practice.

The main research question explored in this chapter was: *In the context of customised production, how do hybrid rescheduling algorithms impact entropic-related complexity?* The following objectives guided this chapter: (a) To identify the typical hybrid rescheduling algorithms used in the context of customised production, and (b) To explain why and how hybrid rescheduling algorithms vary across organisations.

For the research question, it can be seen from the results that *Priority Low* (originated from Company A) generates higher complexity indices than *Utilisation High* (originated from Company B). This can be dependent on the context of the organisations studied, for example production lead times were longer at Company A than Company B. It was also confirmed that the less rescheduling is carried out the lower the complexity, with Company C (SME making customised production) showing a level of complexity management that could serve as a beacon for other companies. As a reminder, Company C embedded flexibility into their manufacturing process by allocating spare capacity in their production schedule.

A counter-intuitive finding was that the entropic-related complexity associated with customised production was lower than that related to commodity production. This could be explained by the fact that commoditised production typically faces tougher competition; the customer could easily switch from one supplier to another and this can happen as late as already placed orders or production set into motion (Sivadasan *et al.,* 2010). By contrast, in the case of customised production, once the customer has made up their mind and agreed to the specification of their product (which had been discussed in order to match its specifications), then the company is protected (or bounded) by contract to provide what has been agreed. This leads to a more stable production schedule in practice.

Regarding the first objective, it could be argued that manufacturing organisations in customised production prefer the two hybrid rescheduling algorithms identified in this chapter: *Priority Low* and *Utilisation High*. However, this is dependent on the companies selected for the case studies, so different companies may use different hybrid rescheduling algorithms. With respect to the second objective, several other factors not included in this chapter could have an influence on the different choices of hybrid rescheduling algorithms by manufacturing companies, production lead times (e.g. weeks or months) or the type of competition (e.g. sheltered or competitive).

Implications for Theory and Practice

This chapter brings a novel and analytically-based insight into the effects of hybrid rescheduling algorithms on entropic-related complexity. An unpredicted result—of customised production leading to lower complexity than mass production—is highlighted and explained. The work presented in this chapter has shown the application of concepts derived from previous research carried out by the authors to real-world manufacturing companies. As stated earlier, the two quadrants of interest to the work presented in this chapter were: "functional product, responsive scheduling" and "innovative product, responsive scheduling" (Huaccho Huatuco, 2003). The former refers to commodity production, whereas the latter refers to customised production.

Organisations that belong to the "functional product, responsive scheduling" quadrant can manage complexity through decision-making and rescheduling. Here the key point is to be responsive to customer requests by adapting the current scheduled quantity. This was observed in previous case studies carried out by the authors (Sivadasan *et al.*, 2010).

Organisations that belong to the "innovative product, responsive scheduling" quadrant, which are made-to-order of customised products, can deal with complexity through spare capacity and rescheduling in order to become responsive. In this chapter it has been found that *Priority Low* and *Utilisation High* are the preferred hybrid rescheduling algorithms used by manufacturing organisations in practice. However, these results should not be taken prescriptively. Each company should identify its main objectives, and design and manage the system consistently with these objectives. The thresholds/ tolerance limits / trade-offs of spare capacity should be analytically investigated and linked with each company's ability to manage the additional complexity, either value- or non-value adding, associated with these decisions. In this line, companies should reserve spare capacity for products at greater risk of mismatch between their supply and demand, since additional spare capacity in the production schedule can be a detriment to internal performance measures and a drain on costs.

Future work should consider the possibility of carrying out extra case studies for customised production to see whether the other rescheduling strategies are used in practice: *Priority High, Utilisation Low* and *Right Shift* in the context of customised production environment. Another future work direction could be to design, set up and run computer simulations, which should include some spare capacity already built-in the Original Schedule (OS), a sign of robust and sound scheduling methods. Then, a comprehensive statistical analysis of results could be performed.

REFERENCES

Bean, J. C., Birge, J. R., Mittenthal, J., & Noon, C. E. (1991). Matchup scheduling with multiple resources, release dates and disruptions. *Operations Research*, *39*(3), 470–483. doi:10.1287/opre.39.3.470

Bierwirth, C., & Mattfeld, D. C. (1999). Production scheduling and rescheduling with genetic algorithms. *Evolutionary Computation*, *7*(1), 1–17. doi:10.1162/evco.1999.7.1.1

Browning, T. R., Deyst, J. J., & Eppinger, S. D. (2002). Adding value in Product Development by Creating Information and Reducing Risk. *IEEE Transactions on Engineering Management*, *49*(4), 443–458. doi:10.1109/TEM.2002.806710

Calinescu, A., Efstathiou, J., Schirn, J., & Bermejo, J. (1998). Applying and Assessing Two Methods for Measuring Complexity in Manufacturing. [JORS]. *The Journal of the Operational Research Society*, *49*(7), 722–733.

Calinescu, A., Efstathiou, J., Sivadasan, S., Schirn, J., & Huaccho Huatuco, L. (2000). Complexity in manufacturing: an information theoretic approach, In I.P. McCarthy & T. Rakotobe-Joel, eds. *Proceedings of the conference on complexity and complex systems in industry*, pp.30-44. Warwick: The University of Warwick.

Cheng, Y. (1998). Hybrid simulation for resolving resource conflicts in train traffic rescheduling. *Computers in Industry*, *35*(3), 233–246. doi:10.1016/S0166-3615(97)00071-7

Da Silveira, G., Borenstein, D., & Fogliatto, F. S. (2001). Mass customization: Literature review and research directions. *International Journal of Production Economics*, *72*(1), 1–13. doi:10.1016/S0925-5273(00)00079-7

Deblaere, F., Demeulemeester, E., & Herroelen, W. (2010). Reactive scheduling in the multi-mode RCPSP. *Computers & Operations Research*. doi:. doi:10.1016/j.cor.2010.01.001

Duray, R., Ward, P. T., Milligan, G. W., & Berry, W. L. (2000). Approaches to mass customisation: configuration and empirical validation. *Journal of Operations Management*, *18*(6), 605–625. doi:10.1016/S0272-6963(00)00043-7

Efstathiou, J. (1996). Anytime heuristic schedule repair in manufacturing industry. *IEE Proceedings. Control Theory and Applications*, *143*(2), 114–124. doi:10.1049/ip-cta:19960277

Efstathiou, J., Tassano, F., Sivadasan, S., Shirazi, R., Alves, J., Frizelle, G., & Calinescu, A. (1999). Information complexity as a driver of emergent phenomena in the business community, *Proceedings of the international workshop on emergent synthesis*, pp. 1-6. Japan: Kobe University.

ElMaragy, H. A. (Ed.). (2009). *Changeable and Reconfigurable Manufacturing Systems*. New York: Springer. doi:10.1007/978-1-84882-067-8

Frizelle, G., & Woodcock, E. (1995). Measuring complexity as an aid to developing operational strategy. *International Journal of Operations & Production Management*, *15*(5), 26–39. doi:10.1108/01443579510083640

Gao, Y., Ding, Y.-S., & Zhang, H.-Y. (2009). Job-shop scheduling considering rescheduling in uncertain dynamic environment. *Proceedings of the 16th International Conference on Management Science & Engineering*. 14th-16th September 2009. Moscow, Russia.

Griffiths, J., & Margetts, D. (2000). Variation in production schedules – implications for both the company and its suppliers. *Journal of Materials Processing Technology*, *103*(1), 155–159. doi:10.1016/S0924-0136(00)00408-8

Gunasekaran, A., & Ngai, E. W. T. (2005). Build-to-order supply chain management: a literature review and framework for development. *Journal of Operations Management*, *23*(5), 423–451. doi:10.1016/j.jom.2004.10.005

Herrmann, J. W. (2001). *Improving manufacturing system performance through rescheduling [online]*. Maryland, University of Maryland. Available from: http://www.isr.umd.edu/~jwh2/papers/rescheduling.html [Accessed 3rd December 2010].

Herrmann, J. W., & Delalio, D. R. (2001). Algorithms for Sheet Metal Nesting. *IEEE Transactions on Robotics and Automation*, *17*(2), 183–190. doi:10.1109/70.928563

Herrmann, J. W., & Pundoor, G. (2002). *Rescheduling frequency and Supply Chain performance*. Institute for Systems Research (ISR). TR 2002-50. 25 p.

Holthaus, O. (1999). Scheduling in job shops with machine breakdowns: an experimental study. *Computers & Industrial Engineering*, *36*, 137–162. doi:10.1016/S0360-8352(99)00006-6

Hornby, A. S. (2000). *Oxford Advanced Learner's Dictionary of Current English*. 6th edition. Wehmeier Sally (Ed). Oxford: Oxford University Press, pp. 29 & 638.

Huaccho Huatuco, L. (2006). *Rescheduling and Complexity for Customised Products. Final report of the Seed corn fund research project. Leeds University Business School* (p. 19). UK: LUBS.

Huaccho Huatuco, L., Burgess, T. F., & Shaw, N. E. (2010). Entropic-related complexity for reengineering a robust supply chain: a case study. *Production Planning & Control.* Advance online publication, 15th July 2010, doi 10.1080/09537281003596185, 12 p.

Huaccho Huatuco, L., Efstathiou, J., Calinescu, A., Sivadasan, S., & Kariuki, S. (2009). Comparing the impact of different rescheduling strategies on the entropic-related complexity of manufacturing systems. *International Journal of Production Research*, *47*(15), 4305–4325. doi:10.1080/00207540701871036

Huaccho Huatuco, L., Efstathiou, J., Sivadasan, S., & Calinescu, A. (2001). The value of dynamic complexity in manufacturing systems. *Proceedings of the International Conference of the Production and Operations Management Society (POMS-Brazil 2001).* 11th -14th August 2001, Escola de Administração de Empresas de São Paulo (Business Administration School of Sao Paulo), Brazil, 180-188.

Huaccho Huatuco, L. D. (2003). *The role of rescheduling in managing manufacturing systems' complexity*. DPhil thesis. Oxford:University of Oxford, pp. 237.

Jain, A. K., & ElMaraghy, H. A. (1997). Production scheduling/rescheduling in flexible manufacturing. *International Journal of Production Research*, *35*(1), 281–289. doi:10.1080/002075497196082

Krajewski, L., Wei, J. C., & Tang, L.-L. (2005). Responding to schedule changes in build-to-order supply chains. *Journal of Operations Management*, *23*(5), 452–469. doi:10.1016/j.jom.2004.10.006

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:10.1037/h0043158

Mintzberg, H. (1988). Generic strategies: Toward a comprehensive framework. *Advances in Strategic Management*, *5*(1), 1–67.

Monostori, L., Kadar, B., Pfeiffer, A., & Karnok, D. (2007). Solution approaches to real-time control of customised mass production. *Annals of the CIRP*, *56*(1), 431–434. doi:10.1016/j.cirp.2007.05.103

Morton, T. E., & Pentico, D. W. (1993). *Heuristic Scheduling Systems: with applications to production systems and project management. Wiley Series in Engineering and Technology Management*. New York: John Wiley and Sons.

New, S., & Ramsay, J. (1997). A critical appraisal of aspects of the lean chain approach. *European Journal of Purchasing & Supply Management*, *3*(2), 93–102. doi:10.1016/S0969-7012(96)00019-6

Ouelhadj, D., & Petrovic, S. (2009). A survey of dynamic scheduling in manufacturing systems. *Journal of Scheduling*, *12*(4), 417–431. doi:10.1007/s10951-008-0090-8

Pfeiffer, A., Kadar, B., & Monostori, L. (2007). Stability-oriented evaluation of rescheduling strategies, by using simulation. *Computers in Industry*, *58*(7), 630–643. doi:10.1016/j.compind.2007.05.009

Pinedo, M. (2008). *Scheduling: Theory, Algorithms, and Systems*, 3rd Edition. New York: Springer Science+Business Media, LLC.

Shannon, C. E. (1949). The mathematical theory of communication. In Shannon, C. E., & Weaver, W. (Eds.), *The mathematical theory of communication* (pp. 3–91). Illinois: University of Illinois Press.

Silva, C. A., Sousa, J. M. C., & Runkler, T. A. (2008). Rescheduling and optimization of logistic processes using GA and ACO. *Engineering Applications of Artificial Intelligence*, *21*(3), 343–352. doi:10.1016/j.engappai.2007.08.006

Sivadasan, S., Efstathiou, J., Calinescu, A., & Huaccho Huatuco, L. (2001).A discussion of the issues of state definition in the entropy-base measure of operational complexity across supplier-customer systems. *Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2001)*, 22nd – 25th July 2001, Orlando, Florida, USA, Vol. II, 227-232.

Sivadasan, S., Efstathiou, J., Calinescu, A., & Huaccho Huatuco, L. (2004). Supply Chain Complexity. In New, S., & Westbrook, R. (Eds.), *Understanding Supply Chains* (pp. 133–163). Oxford: Oxford University Press.

Sivadasan, S., Efstathiou, J., Calinescu, A., & Huaccho Huatuco, L. (2006). Advances on Measuring the Operational Complexity of Supplier-Customer Systems. *European Journal of Operational Research*, *171*(1), 208–226. doi:10.1016/j.ejor.2004.08.032

Sivadasan, S., Efstathiou, J., Frizelle, G., Shirazi, R., & Calinescu, A. (2002). An Information-Theoretic Methodology for Measuring the Operational Complexity of Supplier-Customer Systems. *International Journal of Operations & Production Management*, *22*(1), 80–102. doi:10.1108/01443570210412088

Sivadasan, S., Smart, J., Huaccho Huatuco, L., & Calinescu, A. (2010). Operational Complexity and Supplier-Customer Integration: Case studies insights and complexity rebound. [JORS]. *The Journal of the Operational Research Society*, *61*(12), 1709–1718. doi:10.1057/jors.2009.138

Smart, J. (2009). Book reviews. *International Journal of Production Research*, *47*(6), 1713–1714. doi:10.1080/00207540802316568

Smith, S. F. (2002). Technologies for dynamic scheduling. In Workshop on on-line planning and scheduling of the *6th International Conference on Artificial Intelligence Planning and Scheduling (AIPS'02)*. American Association for Artificial Intelligence / European Association of Excellence in AI Planning (AAAI/PLANET).

Tu, Q., Vonderembse, M. A., & Ragu-Nathan, T. S. (2001). The impact of time-based manufacturing practices on mass customization and value to customer. *Journal of Operations Management*, *19*(2), 201–217. doi:10.1016/S0272-6963(00)00056-5

Vieira, G. E., Herrmann, J. W., & Lin, E. (2003). Rescheduling Manufacturing Systems: A Framework of Strategies, Policies, and Methods. *Journal of Scheduling*, *6*(1), 39–62. doi:10.1023/A:1022235519958

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, *22*(2), 195–219. doi:10.1108/01443570210414329

Wang, L., Pan, Q.-K., Suganthan, P. N., Wang, W.-H., & Wang, Y.-M. (2010). A novel hybrid discrete differential evolution algorithm for blocking flow shop scheduling problems. *Computers & Operations Research*, *37*(3), 509–520. doi:10.1016/j.cor.2008.12.004

Wang, L., Zhang, L., & Zheng, D.-Z. (2006). An effective hybrid genetic algorithm for flow shop scheduling with limited buffers. *Computers & Operations Research*, *33*(10), 2960–2971. doi:10.1016/j.cor.2005.02.028

Yamamoto, M., & Nof, Y. (1985). Scheduling/rescheduling in the manufacturing operating system environment. *International Journal of Production Research*, *23*(4), 705–722. doi:10.1080/00207548508904739

KEY TERMS AND DEFINITIONS

Commodity Production: Commodity production refers to the production of mass produced, standard or off-the-shelf products.

Customised Production: Customised production refers to the production where the customer's preferences for the configuration of a final product are taken into account.

Entropic-Related Complexity: Entropic-related complexity is defined as the expected amount of information required to describe the state of the system (Calinescu *et al*., 2000). It is based on "entropy," which was first proposed in the seminal work by Shannon (1949), and was later adapted to manufacturing systems by Frizelle & Woodcock (1995) and Efstathiou *et al.* (1999). As a result, a number of studies about the entropy measure in the context of manufacturing systems emerged (e.g. Sivadasan *et al.*, 2002, 2004, 2006, 2010 and Huaccho Huatuco *et al.*, 2009). Entropy captures two characteristics of complexity: the variety and uncertainty within the system, corresponding to the structural and operational complexity, respectively (Calinescu *et al*., 2000).

Hybrid Algorithm: Taking the separate definitions of "hybrid" and "algorithm" from the Oxford Dictionary (Hornby, 2000), and combining them, a "Hybrid algorithm" can be defined as the product of mixing two or more sets of rules that must be followed when solving a particular problem. In this chapter the problem refers to rescheduling manufacturing systems.

Non-Value Adding (NVA) Complexity: Non-value-adding (NVA) complexity is the complexity that is not helping the company achieve the satisfaction of its customers (e.g. in the form of late deliveries). This complexity needs to be both monitored and controlled, in order to reduce it and mitigate its possible consequences. For a paper on VA and NVA complexity, please refer to Huaccho Huatuco *et al.* (2001).

Rescheduling: Rescheduling, also known as "predictive-reactive scheduling" (Vieira *et al.*, 2003; Morton & Pentico, 1993), is defined in this chapter as changing the schedule in terms of time, quantity or product specifications in response to disturbances (Huaccho Huatuco *et al*., 2009). Disturbances can be of internal or external nature. Internal disturbances are those that occur within the manufacturing system, such as machine breakdowns; whereas external disturbances are those that occur outside the manufacturing system, such as customer changes (Calinescu *et al.*, 1998). Rescheduling affects the performance of manufacturing organisation(s) (Herrmann, 2001; Hermann & Delalio, 2001; Hermann & Pundoor, 2002). Previous research normally uses traditional measures, such as mean tardiness to evaluate the effectiveness of the rescheduling procedure (e.g. Bean *et al.*, 1991).

Value Adding (VA) Complexity: Value adding (VA) complexity is defined in this chapter as the complexity needed to satisfy the customers (e.g. in the form of "on time" delivery). This complexity needs to be managed too, with the manager judging the level that is suitable to allow into the system. This decision depends on the company, its industry sector and wider context where it operates.

This work was previously published in Hybrid Algorithms for Service, Computing and Manufacturing Systems: Routing and Scheduling Solutions, edited by Jairo R. Montoya-Torres, Angel A. Juan, Luisa Huaccho Huatuco, Javier Faulin and Gloria L. Rodriguez-Verjan, pp. 195-224, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 81
Negotiation Protocol Based on Budget Approach for Adaptive Manufacturing Scheduling

Paolo Renna
University of Basilicata, Italy

Rocco Padalino
University of Basilicata, Italy

ABSTRACT

The research proposed concerns the development of a multi-agent scheduling approach able to support manufacturing systems in different dynamic conditions. The negotiation protocol defined budget approach is based on a financial asset that each part obtains when it is released into the manufacturing system for processing. The part spends the budget to perform the manufacturing operations by the workstations; the virtual market in which part agent and workstation agents coordinate the decentralized system. A fuzzy tool is proposed to assign the budget to each part based on the objectives pursued. A simulation environment based on Rockwell ARENA® platform has been developed in order to test the proposed approach. The simulations are used to compare the proposed approach with classical dynamical scheduling approaches proposed in literature. The results show how the proposed approach leads to better results, and it can be selective among the different priority of the parts.

INTRODUCTION AND MOTIVATION

Nowadays, the competition is played in a market environment characterized by demand fluctuations, new product introduction, reduction of the products life cycle and random disturbances. Some disturbances can be: the failures of manufacturing machines or technological equipment; the importance of the parts can be modified; the objectives of the manufacturing system can change; etc. In this environment, a static scheduling approach leads to reduce drastically the performance of the manufacturing system when exceptions occur. Therefore, the manufacturing systems need

DOI: 10.4018/978-1-4666-1945-6.ch081

to dynamic scheduling to keep a high level of the performance when the conditions change. The dynamic scheduling concerns the possibility to take all decisions based on the current state of the manufacturing system. Ouelhadj and Petrovic (2009) classified the dynamic scheduling in three categories: completely reactive scheduling, predictive–reactive scheduling, and robust pro-active scheduling. (Mehta and Uzsoy 1999; Vieira et al. 2000, 2003; Aytug et al. 2005; Herroelen and Leus 2005). In briefly, the completely reactive scheduling is obtained taking the decisions in real time without a global scheduling in advance. The robust pro-active scheduling is performed building the scheduling in advance trying to capture the exceptions in advance. The predictive-reactive scheduling is a compromise between the two above opposite approaches.

The architecture to support the scheduling approaches in manufacturing systems can be classified in centralized and decentralized systems.

The centralized architectures are not flexible enough to adapt themselves to the dynamism and complexity needed in dynamic manufacturing system conditions.

The main requirements of a dynamic scheduling approach are the following:

- Agility: it is the ability to adapt quickly the scheduling to the dynamic conditions of the manufacturing system;
- Scalability: it is the possibility to add resources into the architecture and the expansion is possible without disrupting the entire architecture established
- Fault tolerance: the architecture can reduce the impact of failures occurred to any parts of the architecture.
- Computational efficiency; the problem of scheduling can be more complex to resolve in a single step.

The above requirements can be obtained by a decentralized architecture (Shen et al., 2001). The Multi Agent Systems are able to support the development of decentralized architecture. A MAS is an artificial system composed of a population of autonomous agents, which cooperate with each other to reach common objectives, while simultaneously each agent pursues individual objectives (Ferber, 1999). Among the several definitions agent, Woodbridge and Jennings (1995): "*an agent is a computer system that is situated in some environment, and that is capable of flexible and autonomous action in this environment in order to meet its design objectives. By flexible we mean that the system must be responsive, proactive, and social*". Multi-agent systems (MAS) have already demonstrated their potential in meeting such complex requirements (Monostori et al., 2006). In manufacturing, such as systems have demonstrated their ability to build up agile and reactive behaviour in various settings like enterprise integration and supply chain management (Swaminathan et al., 1996), dynamic system reconfiguration (Shen et al., 1998), learning manufacturing systems [Monostori et al., 1996), distributed dynamic scheduling (Chiuc and Yih, 1995; Váncza and Márkus, 2000), as well as factory control (Brennan, 1997).

A MAS to pursue a global objective needs to implement a coordination mechanism among the autonomous agents. The most common coordination mechanisms used are the Contract net protocol (Smith 1980), market based, auction based (Siwamogsatham and Saygin, 2004) and game theory (Zhou et al., 2009b).

In this chapter, a coordination mechanism based on market-like approach is proposed. The market like approach works like a classical marketplace where the increase of the demand for a product causes its price to rise, and vice versa, to fall. In the case of the manufacturing systems, the resources become the products and parts are buyers asking for these products. The parts have a budget to spend to perform the manufacturing operations buying a resource. The amount of budget allows the generic part to buy the resource that asks a price lower or equal the budget of the part. Therefore, the level of the budget defines what resources the generic part can buy and the related level of services of the resources (working time, efficiency, quality, etc.). The crucial activities of this coordination approach are the following: how the resources define its price and the budget assigned to the parts when they enter to the manufacturing system. The price of the resources outlines the mechanism of real time scheduling and control of the manufacturing system. The budget assignment to the parts interprets the objectives to pursue in terms of priority of the parts typology, due date, throughput, throughput time, etc.

The aim of this approach is to capture the dynamic conditions of the manufacturing system keeping a high level of performance and able to modify easily the objectives changing only one parameter: the budget assigned to the parts. A fuzzy tool is developed to assign the budget to each part based on the objectives of the manufacturing system. A simulation environment based on Rockwell ARENA® platform has been developed in order to test the proposed approach. In particular, the proposed approach is compared with a classical market like approach proposed in literature in order to highlight the behaviour in very dynamic conditions.

The chapter is organised as it follows. In Section 2 the discussion of the literature review is reported. In Section 3 describes the manufacturing environment and the main principles of our proposed distributed control method in Section 4. In section 5, the experimental environment is outlined. Finally, in section 6 the simulation test results are detailed while conclusions are drawn in section 7.

LITERATURE REVIEW

Many authors have addressed the problem of adaptive scheduling related to job shop with most of the approaches proposed based on heuristics and use dispatching rules.

Jeong and Kim (2005) have used simulation to evaluate the manufacturing system performance and selected an appropriate dispatching rule for scheduling the subsequent period. Chryssolouris and Subramaniam (2001) have proposed a dynamic scheduling approach based on a genetic algorithm (GA). At each event of the manufacturing system (machine breakdown, job arrival, mix change, etc.), the GA was run to select a new scheduling method. Trentesaux et al. (2000) proposed a distributed management system that selected the dispatching rule for an FMS based on the state of the manufacturing system. This research was rather focused on the design of a scheduling system. Maione et al. (2003) presented a MAS with a negotiation protocol using also fuzzy logic. As they found, this soft computing technique led to better performance but required more computational time in order to obtain acceptable results.

Cowling et al. (2001, 2003, 2004) and Ouelhadj et al. (2003a, 2003b) proposed a novel multi-agent architecture for integrated and dynamic scheduling in steel production. Each steel production process is represented by an agent, including the continuous caster agents, the hot strip mill agent, the slab yard agent, and the user agent. The hot strip mill and continuous caster agents perform the robust predictive–reactive scheduling of the hot strip mill and the continuous caster, respectively. Robust predictive–reactive scheduling generates robust predictive–reactive schedules in the presence of real-time events using utility, stability, and robustness measures and a variety of rescheduling heuristics.

Aydin and Fogarty (2004) proposed a parallel implementation of the modular simulated annealing algorithm for classical job-shop scheduling. A multi agent-system based on Java technology is proposed. The empirical results obtained show that the method proposed is scalable and decreases the CPU time compared with other approaches proposed in the literature.

Wonga et al. (2006) proposed an agent-based approach for the dynamic integration of the process planning and scheduling functions. The simulations show that the hybrid-based MAS, with the introduction of supervisory control, is able to provide integrated process plan and job shop scheduling solutions with a better global performance.

Zhou et al. (2009a) proposed an agent-based Decision Support System (DSS) for the effective dynamic scheduling of a Flexible Manufacturing System (FMS). The proposed DSS mainly includes six components: User Interface Agent (UIA), Criteria And Rules Selection Agent (CRSA), Performance Evaluation Agent (PEA), Scheduling Decision Selection Agent (SDSA), FMS database and Scheduling Knowledge Base (SKB). The functions of each component are discussed and a corresponding prototype system is developed. Finally, some examples are used to illustrate the decision process and to study the performance of the FMS under different dynamic disturbances.

Recently, some authors have developed coordination approaches inspired by the behavior of social insects like ants, bees, termites and wasp to. Valckenaers and Van Brussel (2005) presented the design of a holonic manufacturing execution system, that is an instantiation of the PROSA architecture (1998). The control and coordination mechanism is inspired by the ant colony foraging behaviour.

Zhou et al. (2008) proposed an algorithm based on Ant Colony Optimization in a shop floor scenario with three levels of machine utilization, three different processing time distributions, and three different performance measures for intermediate scheduling problems. The experimental results show that ACO outperforms other approaches when the machine utilization or the variation of processing times is not high. The procedure proposed is a centralized approach.

Renna (2009) developed two pheromone approaches for the job shop scheduling problem in a Multi Agent System environment. One is based on the past information of a part (throughput time of the manufacturing cell); and the other on the queue of the manufacturing cell. The proposed approaches are tested in a dynamic environment; the simulation results show how the approach based on the queue of the manufacturing cell performs better when the environment conditions are very dynamic.

Renna (2011) proposed a coordination approach for the multi-agent architecture based on the computation of internal and external indexes of the generic manufacturing cell. Several scenarios were considered: from static to very dynamic conditions for internal and external exceptions of the manufacturing system. The simulation results highlighted that the performance of the proposed approach outperforms the performance of a classical workload approach (benchmark) in all conditions. In the related FMS literature there are quite a few publications that deal with an extremely dynamic environment, where also the manufacturing system itself changes frequently. Most of the researches take a centralized approach and propose decision models that require intensive computations, hence are inadequate to control the job shop in real time. In this chapter, an adaptive scheduling approach is proposed that is capable to operate in a very dynamic environment. The proposed method is based on a budget that each job can spend on purchasing the service of manufacturing resources. The price computation of the cell is based on the evaluation of its local state as well as of the global state of the manufacturing system. The main motivations of our investigations are the following:

- Developing a simple methodology where only one parameter has to be set in order to keep a proper performance level of the manufacturing system

- Reducing the information communication among the agents of the architecture in order to obtain an approach to implement in real case applications
- Evaluates the performance of the proposed approach in a very dynamic environment where internal and external exceptions occur
- Integrating our model into a supply chain or extended enterprise. In fact, the only typology of information exchanged between the shop floor and the upper management levels is the importance of each job type that determines the budget assigned to the job

The principal advantage of our method is the possibility to adapt the scheduling approach to the change of the objective of the production planning level by setting only the parameter of the budget.

MANUFACTURING SYSTEM CONTEXT

The testbed for the proposed negotiation approach will be a production system consisting of a given number of cells. Each cell can perform any kind of manufacturing operations so that the resulting system is generic. The manufacturing system processes jobs, each consisting of operations to be executed sequentially. Jobs belong to various job types.

Scheduling decides about the assignment of a manufacturing cell to the next operations of running jobs, therefore it is a pure dispatching problem. If such decisions are made in a distributed way, the dispatching problem is to be solved through a negotiation between autonomous agents representing cells and jobs in a real time fashion.

A cell agent is associated to each workstation; it is an intelligent entity aimed at scheduling the cell's work with as high efficiency as possible. Moreover, when a new job arrives to the manufacturing system the corresponding job agent is created; it analyses the job's status locating the following operation to be scheduled. Dispatching problems for a given job are handled by negotiation between job and cell agents using a version of the well-proven contract net protocol (Smith, 1980). Accordingly, the negotiation process consists of the following steps:

1. Job agent analyses the job status and determines the job's next technological operation to be executed
2. Job agent sends a message to the cell agents informing them that an operation is pending for processing; afterwards, it remains waiting for the cell agents' answers
3. Each cell agent evaluates the workstation status at negotiation time t, calculates and provides a price
4. Job agent receives evaluations coming from all the cells and builds up an index for evaluating each cell's offer
5. Finally, the job is assigned to the cell that provides the best offer at the time t.

As the reader can notice, the above procedure is a simple auction based negotiation protocol. It defines the environmental relations of the autonomous agents involved in the work, but makes no assumption towards the agents' decision-making mechanisms. This means that the above protocol can be adapted to different objectives and decision mechanisms of the autonomous cell and job agents.

The assumptions of the Job-shop scheduling problem researched in this chapter are the following:

- There are n tasks, and each task consists of m operations.
- Each typology part has been given processing order, processing time and due date.
- Orders for production of different parts arrive randomly.
- Operations cannot be pre-empted.

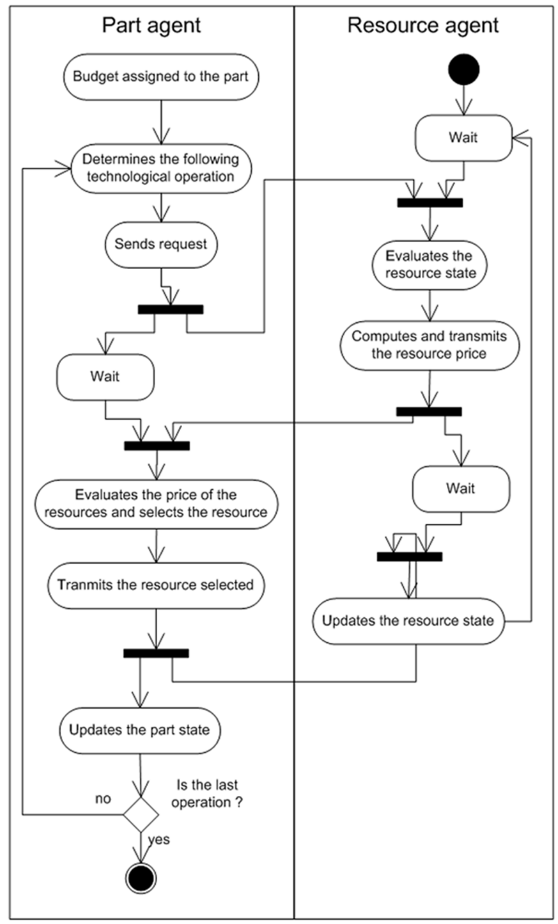
- Each machine can process only one task at once.
- Each machine performs the manufacturing operation with an efficiency, which sets the speed of the operation.
- The queues are managed by the First In First Out policy in order to investigate only the proposed strategy.
- Each machine can breakdown randomly.

In this research, the material handling time is included in the machining time, and the handling resources are always available.

BUDGET APPROACH

Our proposed approach is based on a budget level assigned to each job. The budget is an amount of fictitious money that the job can spend to acquire the manufacturing services. The financial aspects are not considered because we handle problems on the shop floor level where no real transfer of money is made. Essentially, the budget is used to coordinate the multi-agent system. The activity UML diagram in Figure 1 shows the process, with the following activities:

Figure 1. UML activity diagram



1. *Budget assigned to the part*. When a part enters to the manufacturing system, the amount of the budget is assigned to the part. In the following sub-sections is described the fuzzy tool proposed to define the amount of the budget.
2. *Determines the following technological operation*. The part agent determines the technological operation to perform based on the process plan of the part.
3. *Sends request*. The part agent sends a request to all manufacturing resources agent with the information related to the technological operation requested. The part agent waits for the answers by the resources' agent.
4. *Evaluates the resource state*. The resource agent evaluates the processing time of the technological operation requested, the queues of the parts, the failure state, etc.
5. *Computes and transmits the resource price*. The resource agent computes the price to submit to the part agent evaluating the resource state. As described in the following, a high price means high performance of the resource. Then, the resource agent submits the price to the part agent and waits for the part agent answer.
6. *Evaluates the price of the resource and selects the resource*. The part agent collects the price of the resources' agent and selects the price that the part agent can spend (based on

the amount of the budget). In particular, the part agent selects the higher price that can spend in order to buy the better resource. In order to avoid a deadlock, if the budget is not enough to buy any resources, the part is assigned to the worst resource (the resource with the lower price).

7. *Transmits the resource selected*. The part agent transmits the assignation of the part to the resource to the resources' agent.
8. *Updates the resource state*. The resource agent updates the state of the resource in terms of queues and technological operation to perform.
9. *Updates the part state*. The part agent updates the information of the parts and verifies if other technological operations are required. If the part needs another technological operation, the above process is repeated when the part ends the technological operation assigned.

PERFORMANCE INDICES AND PRICE COMPUTATION

The price computed by the resource agent is based on the state of the manufacturing resource.

Details of the price computation are given by the following expressions. The first expression computes a performance about the flow time of the parts, that is between 0 (worst performance) and 1 (better performance, no parts are in the queue).

$$NorFlowTime_m(t) = \frac{Max\Pr ocTime_{Type_{j,k}}}{FlowTime_m(t) + Max\Pr ocTime_{Type_{j,k}}} \quad (1)$$

where m is the index of the resource, $Max\Pr ocTime_{Type_{j,k}}$ is the maximum processing time for the technological operation k of the part of type j.

The effective service time $\left(EWT_j^k(t)\right)$ is obtained by multiplying this value with the resource's efficiency CTC_m, that is minor of one. A greater value of CTC_m, means that the generic resource is able to process the part with higher speed.

$FlowTime_m(t)$ is the expected throughput time of resource m computed by summing up the processing times of the parts waiting in the resource's queue plus the residual service time of the part being just worked in the cell at the negotiation time t. This index is the measure of workload of the manufacturing cell. The index value is one if no parts are in the queue and the resource is in the idle state, and decreases with the increase of parts in the queue.

Then, a Cell Failure Index, *RFI*, is computed as follows:

$$RFI_m(t) = 1 - \frac{FT_m(t)}{t} \quad (2)$$

where $FT_m(t)$ is the total time of failure status of resource until negotiation time period t. *RFI* above is the index of reliability of the manufacturing resource. This index value is one if no failure happens and decreases with the increase of the failure time.

A Resource Processing Time Index, *RPTI*, is computed as

$$RPTI_m(t) = 1 - CTC_m(t) \quad (3)$$

where $CTC_m(t) > 0$ is the efficiency of resource m at time t. The value of $CTC_m(t)$ multiplied by $TaskT_{Type_{j,k}}$ produces the expected working time of the resource. In particular, a lower value of $CTC_m(t)$ leads to lower expected time to manufacturing a generic part.

Then, an Internal Resource Index, *IRI* is computed by the following expression as a combination of the two above indices:

$$IRI_m(t) = \beta \cdot RFI_m(t) + (1-\beta) \cdot RPTI_m(t) \quad (4)$$

Finally, the External Resource Index, *ERI* is determined as follows:

$$ERI_m(t) = NorFlowTime_m(t) \quad (5)$$

The *IRI* is the average of the indices related to the cell, while the *ERI* is the index related to the manufacturing system status.

The so-called Resource Efficiency Index of the manufacturing resource is the following *REI*:

$$REI_m(t) = \alpha \cdot ERI_m(t) + (1-\alpha) \cdot IRI_m(t) \quad (6)$$

where α is a weight between the internal and external efficiencies. Note that the value of *REI* can assume the values in the interval [0,1].

The resource agent computes the price as follows:

$$Price_m(t) = K \cdot REI_m(t), \quad (7)$$

where the constant *K* is the maximum price of the cell.

Therefore, if the value of the *REI* index is one (the state of the resource is the better possible: no parts in the queue, no failures occur and higher speed of processing time), the resource agent asks the higher price K; otherwise the price reduces proportionally to the reduction of the *REI* index.

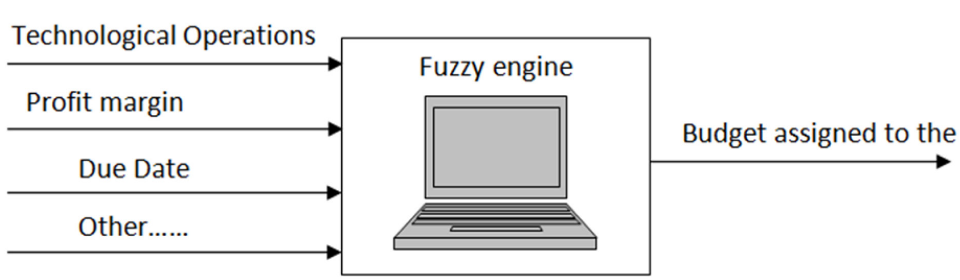
The job agent selects the resource with the highest price required, because high price involves a high level of performance of the manufacturing resource. If the job does not have the sufficient budget to pay the resource, it has to select a resource that requests a price compatible with the budget of the job. If all resources request such a price that the part cannot pay, the part is dispatched to the resource that requests the lowest price (deadlock avoidance). Finally, the part and resource agents update their budget for processing the subsequent operations.

Budget Allocation

A fuzzy approach has been proposed to assign the value of budget to each part type. Figure 2 shows this fuzzy engine whose inputs data are the following:

- Technological operations: the number of operations that the parts have to get processed by the manufacturing system. The increase of this number leads to increase the level of the budget (more resources to buy).
- Profit margin: the level of profit margin related to the product type. The increase of the profit leads to increase the level of the budget (greater importance of the product type).

Figure 2. Inputs and output data of the fuzzy tool



- Due date: a closer due date leads to increase the budget assigned to the part, so that the part can have a chance to get better manufacturing resources in order to minimize the delivery delay.
- Other: other inputs can be easily added to the fuzzy engine to include other characteristics of the part.

The only output of the system is the level of budget allocated to the job.

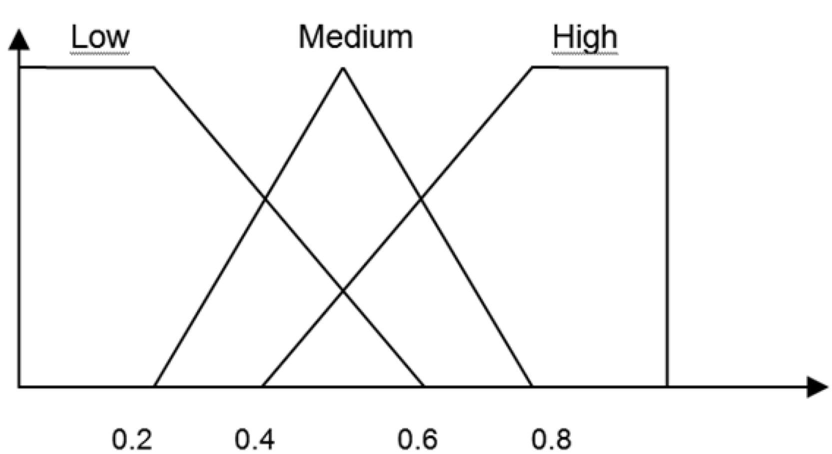
For each input there is defined the following fuzzy set (see Figure 3):

Table 1 shows the value of the inputs of the fuzzy systems for the test case where, as the reader can notice, the technological operations can assume values like 2, 3, 4 and 5. The profit margin of products is classified in three categories: 1 stands for low, 2 for medium, and 3 for high. Finally, the due date input depends on the value of the multiplier of the technological operation time.

Table 1. Fuzzy inputs value

	low	medium	high
Technological operations	2	3 - 4	5
Profit margin	1	2	3
Due date	[3-4]	[2-3)	[1.5-2)

Figure 3. Fuzzy sets



The fuzzy inference rules applied are the following:

no.	rule		budget
1	IF due date is High	THEN	High
2	IF due date is Medium	THEN	Medium
3	IF due date is Low	THEN	Low
4	IF number of technological operations is High	THEN	High
5	IF number of technological operations is Medium T	THEN	Medium
6	IF number of technological operations is Low	THEN	Low
7	IF profit product is High	THEN	High
8	IF product profit is Medium	THEN	Medium
9	IF product profit is Low	THEN	Low

The budget assigned to the part is obtained by activating the above fuzzy rules; among the rules with the same consequent is adopted the rule with higher membership. The rules are arranged by the "maximum" method and the de-fuzzification is done by the centre gravity method. The crisp value *z_out* leads to assign the budget to the jobs by the following expression:

$$Budget = \begin{cases} High & 0.625 \le z\_out \le 1; \\ Medium & 0.325 \le z\_out < 0.625; \\ Low & 0 \le z\_out < 0.325; \end{cases} \tag{8}$$

Alternative Solution Methods

In order to validate the operation of the proposed approach, it has been compared with other on-line scheduling methods; some of them taken from the literature (Jeong and Kim,1998; Smith, 1980). For all these solvers, the negotiation protocol is the same as presented in Section 3. Where they differ is the offer formulation and how the resource's choice is made. The following six methods have been experimented with:

- Budget: The Budget Approach, as described above.
- Budget+DIS: The Budget Approach is combined with the heuristic dispatching rule LowValueFirst: DueDate.
- Slack1: Each cell evaluates Service Time End (STE), the estimated finish time of the task subject to negotiation. STE is calculated in the following way:

$$STE_m(t) := \sum_{i}^{NQ} EWT_j^k(t) + Res_m(t) \quad \forall j \in Queue_m \tag{9}$$

where *NQ* is the number of jobs in the cell's queue, while Res_m is the estimated remaining operational time of the actual running job. The job agent calculates *SlackOp* in the following way:

$$SlackOp := \frac{DueDate}{\sum_{k=k^*}^{Ntask} Max\Pr ocTime_k} \tag{10}$$

where *k* is the task under decision. Job agent choices the cell which has the following minimum difference:

if *SlackOp* > 0, $\min_m (SlackOp - STE_m)$

If *SlackOp* < 0, $\min_m (STE_m)$ (11)

Slack2: As the previous one, but *SlackOP* is divided by 2, to compensate for system delays. Several values are tested and value 2 leads to better performance measures for the test case investigated.

Slack3: *SlackOp* is calculated in a different way:

$$SlackOp := \frac{DueDate}{\sum_{k=k^*}^{Ntask} TaskTime_k} - NQ \cdot MTD - AvTTR \tag{12}$$

where *MTD* is the maximal possible delay in each task execution and *k* is the task under decision; in this model *MTD* = 5. *AvTTR* is the mean failure length.

MinSTE: Each cell evaluates the *STE* (Service Time End), the estimated finish time of the request task as in *Slack1*. The job agent chooses cell *m* corresponding to $\min_m (STE_m)$.

SIMULATION ENVIRONMENT

The manufacturing system processes various parts that arrive according to an exponential distribution. A part type is specified by a sequence of tasks (so-called routing, see Table 2) to be processed and the maximum processing times (*MaxProcTime,* see Table 3) of the tasks. Tasks belong to various task types, each having a fixed routing.

Each *part* $Type_i$ has its $InterArrivalTime_i$, (exponential distribution is used) which can be easily set. Hence, the mix of jobs is determined

Table 2. Routings

Type	Tasks				
Type 1	2	1	3	4	
Type 2	3	2	1		
Type 3	1	4			
Type 4	4	3	1	3	2

Table 3. Maximum process times

Type	$Max\,ProcTime_{Type_j,k}$					TPT
Type 1	20	25	15	18		78
Type 2	30	25	19			74
Type 3	15	27				42
Type 4	20	25	30	15	20	100

TotalProcessTime(*TPT*) is the sum of the *MaxProcTime* for each job type.

stochastically by the *InterArrivalTimes* setting. An example is reported in Table 4. Each part has a *DueDate* that is set when the job arrives at the system. Each *DueDate* is calculated according to the following rule:

$$DueDate_i := CoeffDD_{Type_j} \cdot TPT_{Type_j} \tag{13}$$

The coefficient *CoeffDD* is characteristic to the job type (see Table 5); its actual value is taken from uniform distribution, according to the following rule:

$$CoeffDD_i = unif(CDDmin_i, CDDmax_i).$$

The resources are able to process various task types; in the following table the resource - task type assignment is marked. The system is reconfigurable, so these assignments can be changed. Value 1 indicates that the cell can execute a particular type of task, and 0 otherwise (see Table 6).

Table 4. Inter-arrival times

Type	Type 1	Type 2	Type 3	Type 4
Inter-Arrival Time	15	10	12	18

Table 5. CDDmin and CDDmax

Type	Type 1	Type 2	Type 3	Type 4
$CDDmin_i$	1.5	2	3	2
$CDDmax_i$	2	3	4	4

We assume a real dynamic environment, so we consider a production stage: in each production stage the characteristics of the manufacturing resource CTC_m (see Table 7) and the time between failures TBF_m are modified. So the Expected Working Time *EWT* depends on CTC_m and it is:

$$EWT_j^k(t) = CTC_m(t) \cdot Max\,ProcTime_{Type_j,k} \tag{14}$$

The *TaskDelay* is calculated according to the following rule:

$$TaskDelay := \begin{cases} 0 & P(p < \gamma) \\ unif(1,5) & P(p > \gamma) \quad where\ \gamma = 0.5 \end{cases} \tag{15}$$

p is a random number between 0 and 1 that emulates the probability that an unforeseen event causes a delay of the part during a manufacturing

Table 6. Resource - task type assignment

Resource	1-6	2-7	3-8	4-9	5-10
Task 1	0	1	0	1	1
Task 2	1	1	1	0	0
Task 3	1	0	1	1	0
Task 4	0	1	1	0	1

Table 7. CTCm

Resource	1-6	2-7	3-8	4-9	5-10
Stage 1	0.30	0.60	0.40	0.50	0.40
Stage 2	0.33	0.66	0.44	0.55	0.44
Stage 3	0.36	0.72	0.45	0.60	0.52
Stage 4	0.40	0.80	0.48	0.65	0.58
Stage 5	0.45	0.88	0.52	0.70	0.65

Table 8. TBFm in the stages

Cell	1-6	2-7	3-8	4-9	5-10
Stage 1	12.00	8.00	10.10	9.00	10.00
Stage 2	10.60	6.40	8.40	8.50	9.00
Stage 3	9.30	5.12	6.10	7.40	8.00
Stage 4	7.70	4.10	4.90	6.30	7.00
Stage 5	6.30	3.27	3.40	5.20	6.00

operation. The value γ=0.5 means that the probability of this event is the 50%. The cells may break down, and this failure model is also included in the model. Failure is a stochastic event (see Table 8): each failure happens according to an exponential rate that is $30 \cdot TBF_m$.

Failure Length (FL) is a stochastic variable taken from an exponential distribution: $FL := exp(C)$ where C is a constant. Finally, there are no setup times and transport times, while buffers and the queues can be infinite.

Experimental Plants and Streams

This section describes how a simulation test-bed, model, and case study have been developed in the Arena® simulation environment in order to test the behaviour of the presented approach. All the solvers have been compared with different streams of jobs and experimental plans. The experimental plan is a particular configuration of the manufacturing system. The stream is a sequence of parts arriving into the system. The idea is to compare results in all possible configurations of the manufacturing systems, under different workload and/or production mix conditions. In this section, there are presented 5 different streams of part arrivals and 5 different experimental plants. Hence, there are 25 experiments for each solution method. Each experiment is a combination of a stream and a plant with 50 replications.

- Experimental Plan 1: Plant is deterministic; absence of failure; Number of cells is 10; CTC_m as in Table 7; Cell – Task Type assignment as in Table 6.
- Experimental Plan 2: Number of cells is 10; $CTCm$ as in Table 7; Cell – Task Type assignment as in Table 6; TBF_m as in Table 8; $FL := exp(30)$.
- Experimental Plan 3: Number of cells is 10; CTC_m as in Table 7; Cell – Task Type assignment as in Table 6; TBF_m as in Table 9; $FL := exp(15)$.
- Experimental Plan 4**:** Number of cells is 7; CTC_m as in Table 10; Cell – Task Type assignment as in Table 6; TBF_m as in Table 8; $FL := exp(30)$.
- Experimental Plan 5: Number of cells is 6; CTC_m as in Table 7; Cell – Task Type assignment as in Table 6; TBF_m as in Table 8; $FL := exp(30)$.

In Table 11 there are demonstrated the steam features.

Table 9. TBFm in the stages of Plan 3

Cell	1-6	2-7	3-8	4	5
Stage 1	25	30	20	25	25
Stage 2	24	28	19	23	24
Stage 3	21	25	17	20	21
Stage 4	18	22	15	18	18
Stage 5	15	18	13	16	15

Table 10. CTCm in Plan 4

Cell	1-6	2-7	3	4	5
Stage 1	0.25	0.40	0.30	0.35	0.27
Stage 2	0.26	0.42	0.32	0.36	0.28
Stage 3	0.27	0.45	0.33	0.36	0.29
Stage 4	0.28	0.48	0.34	0.37	0.30
Stage 5	0.29	0.52	0.36	0.38	0.32

Table 11. Streams

Type	CDD_{min}	CDD_{max}	IAT	Task	TPT
Stream1					
1	1.5	2	40	4	78
2	2	3	40	3	74
3	3	4	40	2	42
4	2	4	40	5	100
Stream2					
1	1.3	1.8	30	4	78
2	1.5	1.9	30	3	74
3	1.3	1.7	30	2	42
4	1.6	1.9	30	5	100
Stream3					
1	1.3	1.8	10	4	78
4	1.6	1.9	15	5	100
Stream4					
1	3	5	25	4	78
2	4	6	15	3	74
3	3	4	20	2	42
4	2.5	5	20	5	100
Stream5					
2	4	6	10	3	110
3	3	4	15	2	70

SIMULATION RESULTS

The criteria used to evaluate the solvers are the Flow Time of the parts, percentage of parts in Delayed, Tardiness, Maximum value of Tardiness, Work-in-Process (WIP), and Throughput (see Table 12).

Initially, it has been conducted a sensitivity analysis of the parameters α and β together. The analysis of the simulation results highlight that the range of α and β that assures the better performance is [0.7-0.85];

In this range, the performance measures don't change significantly. Therefore, the parameters are set to the same value of 0.85 for the simulation experiments conducted.

In particular, it can be observed that the best two solvers are *MinSTE* and *Budget*. To improve these solvers it is necessary to combine them with a dispatching rule Earliest Due Date (EDD). In fact, *Budget+EDD* has the best performance crossing both plants and streams. In the following table, there are summarized the performances of all solvers compared to the *Budget* solver. The value is the percentage (%) difference between the solver written in the head cell and *Budget* solver. Except for the throughput, *Budget* approach improves all performance measures of *MinSTE*. The best improvement is in the percentage of delayed of jobs, while as regards throughput *MinSTE* is better than *Budget*. It is important to underline the *Budget* and *MinSTE* seem to work in the same way, independently from Streams or Plants instances. In fact, while *MinSTE* tries to find the faster cell to perform the subsequent task, *Budget* looks for the most efficient cell.

In most of the cases, the fastest cell and the most efficient one coincide; the different performances of these two solvers are thus due to only a small percentage of choices during one simulation. Performances of other solvers are very far from the aforementioned. This is because they are not able to predict the failure or delay of the system. In fact, their performances are acceptable only in the deterministic plant, even if always worse than *MinSTE* and *Budget*. The worst solver is *Slack1*

Table 12. Solver comparison results

Cell	Min STE	Slack 1	Slack 2	Slack 3	Budget + EDD
% Delay Jobs	6.04	1987	211	365	-10.6
Flow Time	1.33	171	87.2	132	-7.72
WIP	0.68	164	84.2	127	-8.39
Throughput	0.38	-2.91	-1.04	-1.72	4.13
Tardiness	4.96	64.2	29.3	30.7	-14.9
Max Tardiness	3.15	143	55.3	65.5	-17.5

because it makes no attempt to forecast delays or failures. However, the forecast of delays or failures is not always so useful, which is demonstrated by the case of *Slack2*. It happens that *Slack2*, where the forecast is made by a heuristic division of the *SlackOp* index, is better than *Slack3*, in which delays and failure forecasts are explicit.

The above simulations have been conducted with the same level of budget assigned to each job, in order to match up to other approaches in the same environment. Moreover, the value of budget is enough high (defined infinity) to avoid the case in which the part doesn't have a sufficient budget to acquire the best cell. The motivation is to compare the proposed approach with the other methodologies in same operational conditions and evaluate the difference of performance measures due only to budget approach.

Further simulations have been conducted in order to test the fuzzy tool that allocates budget to the jobs.

Therefore, three job types have been hypothesized, with three different levels of budget (low, medium and high). The price of each manufacturing cell is computed as in (7) where k=100.

The low value is always 20, while medium and high values change between 20 and 100 with a step of 10. Moreover, the simulations have been conducted with three values of inter-arrival time and two dimensions of the manufacturing systems. The experimental classes are reported in Table 13.

The 25 experimental classes of budget values are simulated for the values of inter-arrival time and number of manufacturing cell reported in Table 14.

The total number of experimental classes simulated is 150.

The simulation results are compared (basis of comparison) with the case of infinity budget assigned to each part typology (parts with same importance). The performance measures investigated are:

Table 13. Budget values combination

Exp. No	Budget – medium level value	Budget- high level value
1	20	20
2	20	40
3	20	60
4	20	80
5	20	100
6	40	20
7	40	40
8	40	60
9	40	80
10	40	100
11	60	20
12	60	40
13	60	60
14	60	80
15	60	100
16	80	20
17	80	40
18	80	60
19	80	80
20	80	100
21	100	20
22	100	40
23	100	60
24	100	80
25	100	100

Table 14. Inter-arrival and manufacturing cell classes

Exp. No	Inter-arrival time	Manufacturing cells
1	13	4
2	16	4
3	19	4
4	5	8
5	7.5	8
6	10	8

- flow time is the average throughput time of the parts.
- tardiness is the sum of the time in delay of the parts.
- number of jobs manufactured in delay is the total parts that are in delay.

Parts with three different importance level enter the manufacturing system, the performance are collected and reported for each importance level. The Figures 4, 5 and 6 report the performance measures for each product importance computing the average values over the inter-arrival time and number of manufacturing cells. There are reported the more significant results.

From the analysis of the aforementioned figures, the following issues can be drawn:

- The products with a high level can gain advantages only if the level of the budget assigned can be the maximum value possible: 100. Moreover, the improvement is better when the weight of medium importance product is the minimum value possible: 20. The advantages are the same for all three performance measures.
- In case of weight of 100 for high importance products, the products with a medium importance level have a very low difference between the weight of 20 and 40. The performance measure of parts in delay is the performance with minor deterioration. Also, the products with low importance have the same trend of the medium importance products.

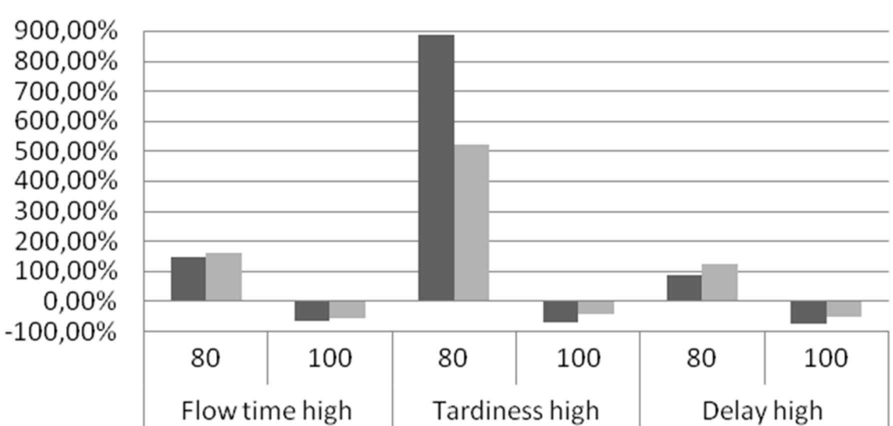
The drastically reduction of flow time of the high important products is obtained with an increment of the flow time of the medium important products about 25 times, while for the low importance products the increment is about 36 times.

Moreover, the products with medium and low importance have a trend of increment of the number of jobs in delay minor than the tardiness performance.

The Figures 7, 8, and 9 report the performance measures for each product importance computing the average value highlighting the manufacturing system dimension (4 an 8) and the inter-arrival time (low, medium and high).

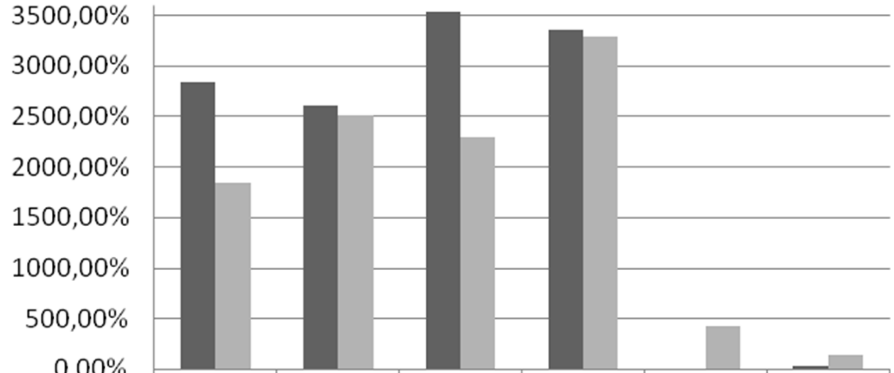
From the analysis of the figures, the following issues can be drawn:

Figure 4. Parts with high importance



	Flow time high 80	Flow time high 100	Tardiness high 80	Tardiness high 100	Delay high 80	Delay high 100
20	146,88%	-67,01%	887,72%	-70,30%	86,51%	-73,89%
40	161,23%	-54,05%	520,83%	-43,97%	123,88%	-52,64%

Figure 5. Parts with medium importance

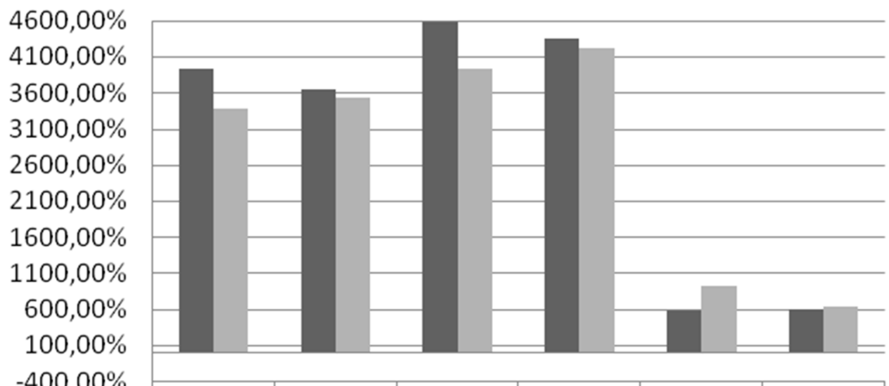


	Flow time medium		Tardiness medium		delay medium	
	80	100	80	100	80	100
20	2837,24	2604,51	3536,95	3350,87	6,04%	32,05%
40	1844,75	2508,31	2295,30	3292,61	426,23%	143,31%

- The performance measures improve when the manufacturing dimension increases from four manufacturing cells to eight.
- From the point of view of manufacturing congestion, the performance measures are better when the inter-arrival is low, therefore with a high congestion level.

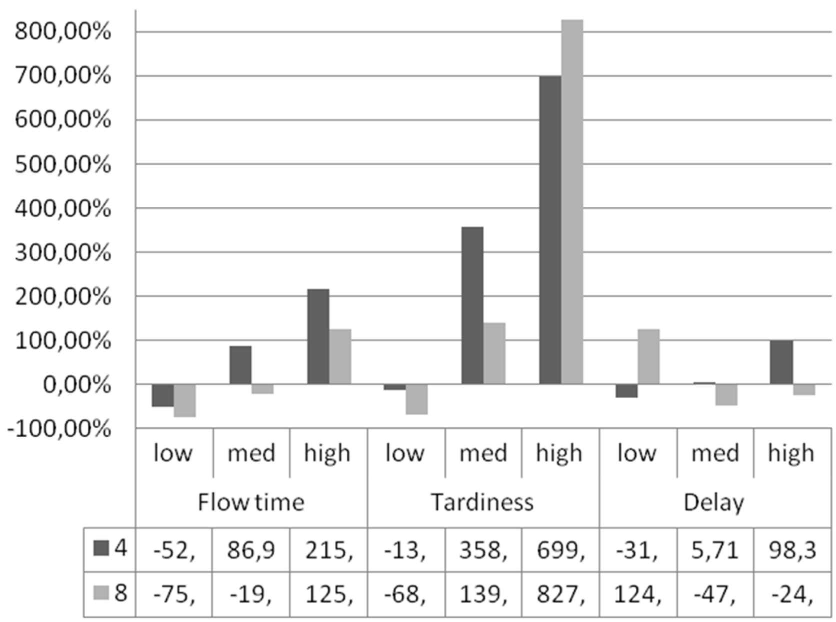
Though, the performance of the jobs with medium and low importance gets worse. The fuzzy

Figure 6. Parts with low importance



	Flow time low		Tardiness low		delay low	
	80	100	80	100	80	100
20	3930,80	3647,31	4579,02	4346,12	574,31%	604,17%
40	3380,03	3543,46	3941,34	4228,26	918,54%	639,93%

Figure 7. Parts with high importance



inference enables assigning a budget to each job correlating to the importance of the job. It can be adapted to the change of the importance of the jobs in real time and this leads to a selective and adaptable. Therefore, the fuzzy reasoning leads to a selective approach

Figure 8. Parts with medium importance

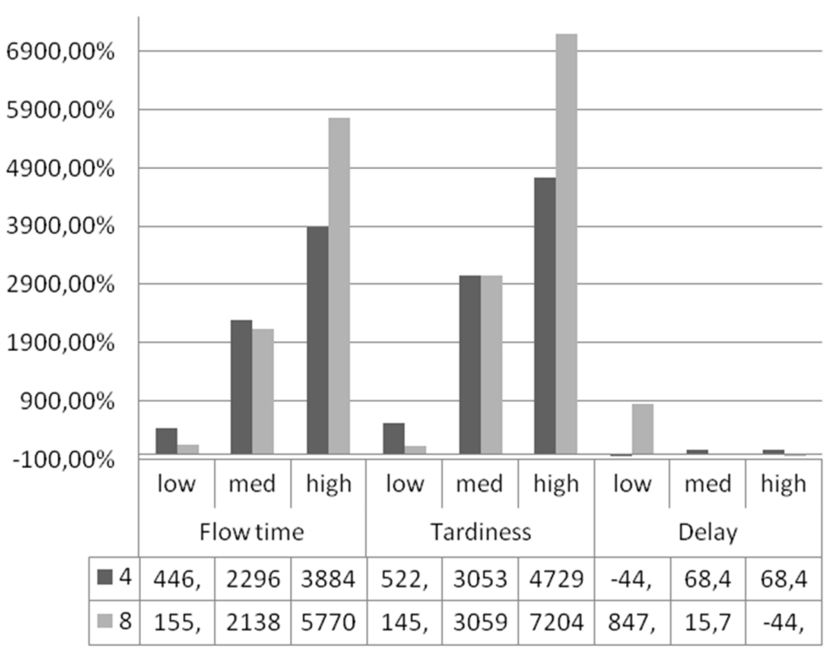
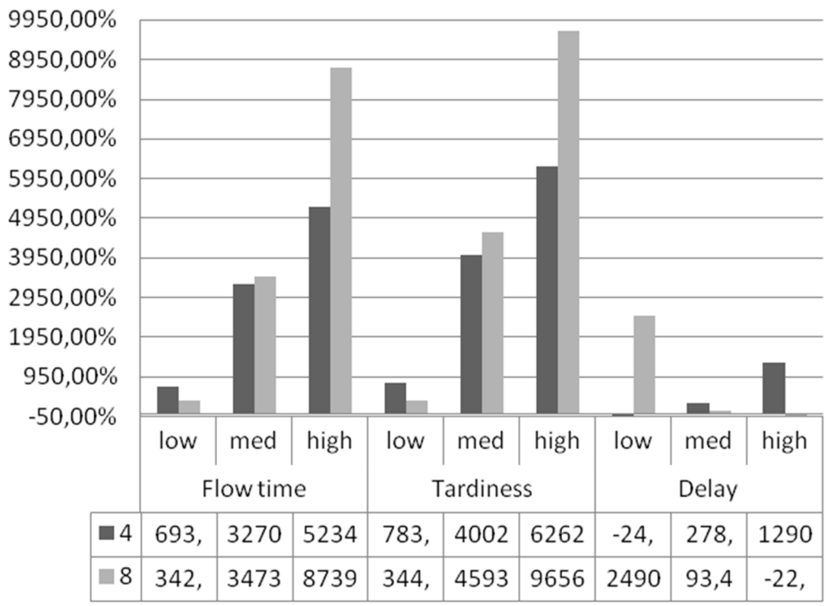


Figure 9. Parts with low importance



	Flow time low	Flow time med	Flow time high	Tardiness low	Tardiness med	Tardiness high	Delay low	Delay med	Delay high
4	693,	3270	5234	783,	4002	6262	-24,	278,	1290
8	342,	3473	8739	344,	4593	9656	2490	93,4	-22,

From the analysis of the simulation results, the following issues can be drawn:

- The budget approach proposed allows to discriminate the performance level by setting the importance of the jobs. The importance of the jobs is set by a fuzzy tool, therefore, defining qualitative parameters.
- The better performance for the high importance jobs is obtained when the configuration of the budget assigned to the part types is the following: HIGH=100, MEDIUM=20 and LOW=20. Then, the fuzzy tool works better when the classes are two.
- The performance measures are better when the congestion level is high (low inter-arrival) and the number of the manufacturing cells increases from four to eight.
- Finally, the extremely worst performance for medium and low importance performance have to be understood as a reduction compared to the case when all the parts have the same importance.

The fuzzy tool proposed allows to modify the performance level of different classes of job importance by setting only one parameter: budget assigned to the parts.

Moreover, the simulation results can be used to learn the opportune level of budget to assign at each typology part when the manufacturing conditions change.

The development of a hybrid approach, in which the proposed approach is activated when the congestion level is high and the selection of the products' importance, is necessary.

CONCLUSION

This chapter proposed an innovative scheduling approach for flexible manufacturing systems working in dynamic environments. Performance measures based on internal and external indices of efficiency have been elaborated. The negotiation protocol has been tested under different conditions of manufacturing system workload, with various numbers of cells and dynamic cell characteristics. The results show how the *Budget* approach leads to better performance, especially if it is combined with appropriate dispatching rules. Merits of the novel method are more characteristic when the manufacturing system is more dynamic. The *Budget* approach was also compared to other dynamic scheduling methods proposed in the literature. In particular, the *Budget* solver has performance much better than other solvers that try to estimate the next task slack because it considers also the possible failure state of cells; on the other hand, *Budget* is better than the *MinSTE* solver just because it is able to select dynamically the more efficient cell, not simply the fastest one. All in all, it has a high level of adaptivity without deteriorating performance. Overall system performance depends only on one intuitive parameter, the budget of the jobs. Several characteristics that define the priority of a part typology are combined by a fuzzy tool in an only parameter: the budget value of the part. The coordination approach is very simple, because the only parameter to set is the budget of the part. Moreover, if a new characteristic has to be added or modified to define the priority of the part, the coordination approach doesn't change, only the fuzzy tool needs to be modified to change the priority of the parts.

Hence, the overall system is well controllable and the approach can be extended towards production networks where the key to common competitiveness and efficiency is just the sharing financial risks and benefits. Future developments concern the application of the proposed approach in a real case study in order o validates the results obtained by the simulation experiments. Moreover, artificial intelligence approaches can be developed to learn the opportune level of budget. In particular, generic algorithms and neural networks can be integrated to learn and select the level of appropriate budget.

REFERENCES

Aydin, M. E., & Fogarty, T. C. (2004). A simulated annealing algorithm for multi-agent systems: A job shop scheduling application. *Journal of Intelligent Manufacturing*, *15*(6), 805–814. doi:10.1023/B:JIMS.0000042665.10086.cf

Aytug, H., Lawley, M. A., McKay, K., Mohan, S., & Uzsoy, R. (2005). Executing production schedules in the face of uncertainties: A review and some future directions. *European Journal of Operational Research*, *161*(1), 86–110. doi:10.1016/j.ejor.2003.08.027

Brennan, R. W., Balasubramanian, S., & Norrie, D. H. (1997). *Dynamic control architecture for advanced manufacturing system.* International Conference on Intelligent Systems for Advanced Manufacturing, Pittsburgh, PA, USA.

Chiuc, C., & Yih, Y. (1995). A learning based methodology for dynamic scheduling in distributed manufacturing systems. *International Journal of Production Research*, *33*(11), 3217–3232. doi:10.1080/00207549508904870

Chryssolouris, G., & Subramaniam, V. (2001). Dynamic scheduling of manufacturing job shops using genetic algorithms. *Journal of Intelligent Manufacturing*, *12*, 281–293. doi:10.1023/A:1011253011638

Cowling, P. I., & Johansson, M. (2002). Using real time information for effective dynamic scheduling. *European Journal of Operational Research*, *139*(2), 230–244. doi:10.1016/S0377-2217(01)00355-1

Cowling, P. I., Ouelhadj, D., & Petrovic, S. (2001). A multi agent architecture for dynamic scheduling of steel hot rolling. In *Proceedings of the third international ICSC world manufacturing congress*, (pp. 104–111). Rochester, NY, USA.

Cowling, P. I., Ouelhadj, D., & Petrovic, S. (2003). A multi-agent architecture for dynamic scheduling of steel hot rolling. *Journal of Intelligent Manufacturing*, *14*, 457–470. doi:10.1023/A:1025701325275

Ferber, J. (1999). *Multi-agent systems: An introduction to distributed artificial intelligence*. London, UK: Addison-Wesley.

Guanghui, Z., Jiang, P., & Huang, G. Q. (2009b). A game-theory approach for job scheduling in networked manufacturing. *International Journal of Advanced Manufacturing Technology*, *41*(9-10), 972–985. doi:10.1007/s00170-008-1539-9

Herroelen, W., & Leus, R. (2005). Project scheduling under uncertainty: Survey and research potentials. *European Journal of Operational Research*, *165*(2), 289–306. doi:10.1016/j.ejor.2004.04.002

Jeong, K.-C., & Kim, Y.-D. (1998). A real-time scheduling mechanism for a flexible manufacturing system: Using simulation and dispatching rules. *International Journal of Production Research*, *36*(9), 2609–2626. doi:10.1080/002075498192733

Jeong, S. J., Lim, S. J., & Kim, K. S. (2005). Hybrid approach to production scheduling using genetic algorithm and simulation. *International Journal of Advanced Manufacturing Technology*, *28*(1-2), 126–136.

Maione, G., & Naso, D. (2003). A soft computing approach for task contracting in multi agent manufacturing control. *Computers in Industry*, *52*, 199–219. doi:10.1016/S0166-3615(03)00127-1

Mehta, S. V., & Uzsoy, R. (1999). Predictable scheduling of a single machine subject to breakdowns. *International Journal of Computer Integrated Manufacturing*, *12*(1), 15–38. doi:10.1080/095119299130443

Monostori, L., Márkus, A., Van Brussel, H., & Westkämper, E. (1996). Machine learning approaches to manufacturing. *Annals of the CIRP*, *45*(2), 675–712.

Monostori, L., Váncza, J., & Kumara, S. R. T. (2006). Agent-based systems for manufacturing. *CIRP Annals - Manufacturing Technology, 55*(2), 697-720.

Ouelhadj, D., Cowling, P. I., & Petrovic, S. (2003a). Contract net protocol for cooperative optimisation and dynamic scheduling of steel production. In Ibraham, A., Franke, K., & Koppen, M. (Eds.), *Intelligent systems design and applications* (pp. 457–470). Berlin, Germany: Springer.

Ouelhadj, D., Cowling, P. I., & Petrovic, S. (2003b). Utility and stability measures for agent-based dynamic scheduling of steel continuous casting. In *Proceedings of the IEEE International Conference on Robotics and Automation* (175–180). Taipei, Taiwan.

Ouelhadj, D., & Petrovic, S. (2009). A survey of dynamic scheduling in manufacturing systems. *Journal of Scheduling*, *12*, 417–431. doi:10.1007/s10951-008-0090-8

Renna, P. (2009). A performance comparison between efficiency and pheromone approaches in dynamic manufacturing scheduling. In Chiong, R. (Ed.), *Intelligent systems for automated learning and adaptation* (pp. 273–298). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-798-0.ch012

Renna, P. (2011). Multi-agent based scheduling in manufacturing cells in a dynamic environment. *International Journal of Production Research*, *49*(5), 1285–1301. doi:10.1080/00207543.2010.518736

Shaw, M. J. (1988). Dynamic scheduling in cellular manufacturing systems: A framework for network decision making. *Journal of Manufacturing Systems*, *13*, 13–94.

Shen, W., Norrie, D. H., & Barthes, J. P. A. (2001). *Multi-agent systems for concurrent intelligent design and manufacturing*. London, UK: Taylor & Francis.

Shen, W., Norrie, D. H., & Xue, D. (1998). *An agent-based manufacturing enterprise infrastructure for distributed integrated intelligent manufacturing system.* In PAAM'98.

Siwamogsatham, V., & Saygin, C. (2004). Auction-based distributed scheduling and control scheme for flexible manufacturing systems. *International Journal of Production Research*, *42*, 547–572. doi:10.1080/00207540310001613683

Smith, R. G. (1980). The contract net protocol: High-level communication and control in distributed problem solving. *IEEE Transactions on Computers C*, *29*(12), 1104–1113. doi:10.1109/TC.1980.1675516

Swaminathan, J. M., Smith, S. F., & Sahed, N. M. (1996). *A multi-agent framework for supply chain dynamics*. NSF Research Planning Workshop on AI & Manufacturing.

Trentesaux, D., Pesin, P., & Tahon, C. (2000). Distributed artificial intelligence for FMS scheduling control and design support. *Journal of Intelligent Manufacturing*, *11*, 573–589. doi:10.1023/A:1026556507109

Valckenaers, P., & Van Brussel, H. (2005). Holonic manufacturing execution systems. *CIRP Annals - Manufacturing Technology, 54*(1), 427-432.

Van Brussel, H., Wyns, J., Valckenaers, P., Bongaerts, L., & Peeters, P. (1998). Reference architecture for holonic manufacturing systems: PROSA. *Computers in Industry*, *37*, 255–274. doi:10.1016/S0166-3615(98)00102-X

Váncza, J., & Márkus, A. (2000). An agent model for incentive-based production scheduling. *Computers in Industry*, *43*, 173–187. doi:10.1016/S0166-3615(00)00066-X

Vieira, G. E., Hermann, J. W., & Lin, E. (2003). Rescheduling manufacturing systems: A framework of strategies, policies and methods. *Journal of Scheduling*, *6*(1), 36–92. doi:10.1023/A:1022235519958

Vieira, G. E., Hermann, J. W., & Lin, E. (2000). Analytical models to predict the performance of a single machine system under periodic and event-driven rescheduling strategies. *International Journal of Production Research*, *38*(8), 1899–1915. doi:10.1080/002075400188654

Wonga, T. N., Leunga, C. W., Maka, K. L., & Fungb, R. Y. K. (2006). Dynamic shopfloor scheduling in multi-agent manufacturing systems. *Expert Systems with Applications*, *31*(3), 486–494. doi:10.1016/j.eswa.2005.09.073

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, *10*(2), 115–152. doi:10.1017/S0269888900008122

Zhou, B., Wang, S., & Xi, L. (2008). Agent-based decision support system for dynamic scheduling of a flexible manufacturing system. *International Journal of Computer Applications in Technology*, *32*(1), 47–62. doi:10.1504/IJCAT.2008.019489

Zhou, R., Nee, A. Y. C., & Lee, H. P. (2009a). Performance of an ant colony optimisation algorithm in dynamic job shop scheduling problems. *International Journal of Production Research*, *47*(11), 2903–2920. doi:10.1080/00207540701644219

KEY TERMS AND DEFINITIONS

Agile Manufacturing System: It is a manufacturing system in which is the organization that has created the processes, tools, and training to enable it to respond quickly to customer needs and market changes while still controlling costs and quality.

Coordination: It is the mechanism how the agent exchange information and cooperate to purse a common objective.

Discrete Event Simulation: It emulates a real system by a chronological sequence of events. Each event occurs at an instant in time and marks a change of state in the system.

Distributed Approach: It is a system where independent cooperate to pursue an objective without centralized information.

Fuzzy Logic: It is a logic derived from fuzzy set theory to deal with reasoning that is fluid or approximate rather than fixed and exact.

Multi-Agent System: It is a system composed of multiple interacting intelligent agents.

Scheduling: It is a decision making process to allocate resources to tasks over a given time periods optimizing one or more objectives.

This work was previously published in Customer-Oriented Global Supply Chains: Concepts for Effective Management, edited by Ephrem Eyob and Edem G. Tetteh, pp. 35-58, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 82
Research Profiles:
Prolegomena to a New Perspective on Innovation Management

Gretchen Jordan
Sandia National Laboratories, USA

Jonathon Mote
Southern Illinois University, USA

Jerald Hage
University of Maryland, USA

ABSTRACT

Despite the increasing importance of the management of research for innovation, the range of differences among types of research, as well as projects and programs, is not adequately captured in current theories of either project or organizational innovation. This chapter offers preliminary discussions for a new perspective about alternative styles of management for different types of research, whether basic, applied, product development, manufacturing, quality control or marketing. Based on these discussions, the chapter proposes a framework for a new perspective of innovation management, called Research Profiles, which is derived from a literature review and extensive field research. This new perspective delineates four research profiles on the basis of two dimensions of research objectives and two dimensions of research tasks. In matching the research objectives and tasks, we identify inherent dilemmas that managers must address and this developing perspective suggests some appropriate research management approaches.

INTRODUCTION

Despite the central importance of scientific and technological research, including product development, for national competitiveness and security, at present there is not an adequate theory about the appropriate managerial styles needed to address alternative kinds of research objectives at the research project, program or inter-organizational level. Organizational innovation theory stemming from Burns and Stalker (1961) typically focuses on the entire organization and, we would suggest, one organizational model (the organic organization), rather than recognizing the existence of differ-

DOI: 10.4018/978-1-4666-1945-6.ch082

ent kinds of research work. More critically, the organic model does not include either the concept of complexity (Brown & Eisenhardt, 1995; Hage, 1999) or external networks of expertise, which are precisely the ones that are increasingly important in the growth of knowledge network communities (Mohrman, Galbraith, & Monge, 2004; Shinn, 2002), and the spread of inter-organizational relationships (Alter & Hage, 1993; Hagedoorn & Duysters, 2002; Powell, 1998; Powell, Koput, & Smith-Doerr, 1996; Van De Ven & Polley, 1992). Indeed, in the organizational innovation literature there is only one study that examines the structure and performance of research laboratories and it does not include external relationships of any kind (Hull, 1988).

Although we are beginning to see an increasing number of studies of research labs (Brown, 1997; Joly & Mangematin, 1996; Jordan, Streit, & Matiasek, 2003; Menke, 1997), inter-organizational alliances (Gomes-Casseres, 1996) and a few studies of research consortia (Browing, Beyer, and Shelter, 1995), the fact remains that none of these studies have connected the measurement of scientific and technological research objectives, to the nature of the research tasks and their appropriate managerial styles. The research literatures cited above stand largely in isolation, often ignoring other kinds of research work. Specifically, the level of the project is overlooked, which is a smaller unit than the organization, the whole organization, and inter-organizational networks of various kinds. Indeed, what makes a proposed theory of management styles necessary is the considerable range in the ways scientific and technology research is organized. While many small research projects funded by the National Science Foundation (NSF) and the National Institutes of Health (NIH), such as those found in academia, tend to be the standard structure, a considerable amount of research is conducted in large-scale organizations and programs, such as mission agencies like the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA), as well as large scale inter-organizational research programs such as the Human Genome Project. For the same reason, the new and growing literature on projects (Brown and Eisenstadt, 1995) overlooks what might be called "Big Science" as represented in the research conducted at the large national and international laboratories such as Argonne in the US and CERN in Switzerland.

Further, as Clarke (2002) has discussed in comprehensive detail, the management of a large number of researchers is very different from the typical management issues involved in contemporary firms or public bureaucracies. Among other differences are the oft-cited assertions that researchers are more motivated by intellectual curiosity than monetary compensation, the longer and more uncertain time horizons for successful objectives, and, perhaps most importantly, work that is seldom standardized and difficult to evaluate.

While a theory about the diversity of research management styles would necessarily differ from more general theories of organizations, the logic in the construction of our perspective is basically the same. First, one must specify particular kinds of research objectives and identify the potential trade-offs. Then one must also distinguish different kinds of research work and tasks. Finally, the management styles appropriate for the linking of the typology of research tasks with a typology of research objectives at the levels of project, program and inter-organization networks have to be determined. In this chapter, we present our argument for a diversity of research management styles in three sections. First, we provide a more detailed justification of the need for our perspective on of management styles. Second, we specify a typology of research work and a typology of research objectives and provide a theoretical linkage between the two. Finally, we offer our proposed view of Research Profiles and discuss the managerial styles necessitated by the kinds of management challenges that are presented in each profile's combination of research tasks and research objectives.

BACKGROUND: THE NEED FOR A NEW PERSPECTIVE OF RESEARCH MANAGEMENT STYLES

Essentially, we identify two lines of argument for justifying a new perspective of management styles. We take as a given that the central importance of scientific and technological research, not only for competitiveness but many other national goals, speaks to the need for greater attention. First, the closest appropriate specialty, namely organizational innovation research, needs to be altered in the light of new theoretical and conceptual developments (see Hage, 1999). Second, the literature on scientific and technological research is missing a conceptual apparatus that would allow for the accumulation of findings across a myriad of studies.

The first argument is based on the fact that most current and past innovation research continues to be dominated by Burns and Stalker's (1961) seminal work on the organic model. In many ways, this model has become outdated. For example, the model is largely focused on new product development by engineers, rather than scientific and technological research. This model also largely ignores the concept of complexity, which has proven to be fundamental when combined with the organic model and the strategy of risk-taking in explaining differences in innovation rates (Hage, 1999). Finally, as argued previously, the organic model is not designed to handle the varying sizes of units in which research is conducted, from the very small project within an organization to the very large mission agencies such as NOAA and NASA or a research consortium such as SEMATECH (Browning, 1995) or various strategic alliances (Gomes-Casseres, 1996).

More recently, the importance of both the organizational learning perspective (Grant, 1996; Kim & Wilemon, 2007) and the theory of idea innovation networks (Hage & Hollingsworth, 2000; Kline & Rosenberg, 1986) are implicitly calling attention to the importance of research as a mechanism for learning. Further, these perspectives suggest the importance of inter-organizational networks that connect two or more of the six arenas of research: basic, applied research, product development, manufacturing research, quality research, and commercialization research. In fact, the latter literature necessitates a considerable revision of the organic model, specifically to allow for external relationships.

In short, the organic model, and organizational innovation research in general, needs to be linked in various ways to the research project literature (Brown & Eisenhardt, 1995; Hobday, 1998; Shenhar, 2001) and the abundant literature on inter-organizational relationships cited above. One of the reasons for this is the dramatic changes that have occurred because of the explosive growth and globalization of R&D and the way in which science and technology are evolving (Hage & Hollingsworth, 2000). Furthermore, the linkages between arenas (for example basic and applied) of research are usually at the project/program level and therefore research projects and the problems of how to manage them should be the analytical focus.

The second line of reasoning that justifies the need for a theory of research management styles is that much of the social studies of science literature have emphasized concrete categories rather than general dimensions that would allow one to accumulate evidence about how best to manage research. For example, the outcomes of research are typically measured in concrete categories, such as papers, patents, peer review assessments, and citations. As we discuss below, we propose a reconceptualization of these ideas so that they can be used as general dimensions, such as the degree of radicalness of the outcome of the research or the scope of the research outcome, that can be applied to basic or applied research, product development research, manufacturing research, and so on. Similarly, research work is often described by such terms as scientific and technological research or the specific content, i.e. physics, chemistry, or biology. Again, there is a need for general dimensions of research tasks that can be applied across these various situations such as the degree of complexity or diversity in the research team and the size of the research program or inter-organizational network. Until these general dimensions are developed and linked, it is not possible to identify the correct managerial practices for linking the nature of the research task with the kind of desired outcome. In summary, the increasing importance of research and these two lines of reasoning provide a compelling case for the importance of our perspective of Research Profiles based on tasks, objectives, as well managerial styles and managerial challenges.

THE BASIC DIMENSIONS FOR RESEARCH MANAGEMENT STYLES

What characteristics might one desire in a categorization of research management styles? Ideally, one should encompass dimensions that tap into the fundamental dilemmas, tensions and problems of conducting research. Since our interest is in making meaningful distinctions, we want to isolate multiple dimensions of research objectives and tasks. Another concern is to connect choices about objectives as much as possible to the existing literature and, in particular, to the literature on innovation.

To provide new insights about research management styles across a wide range of different kinds of research projects, programs and inter-organizational networks, we conducted an inductively-based exploratory study, funded by the U.S. Department of Energy (DOE), to help identify the critical factors facing the research workers and managers. Utilizing the Competing Values Framework as an organizing framework for understanding the pursuit of research (Cameron & Quinn, 1998; Quinn & Rohrbaugh, 1983), a number of focus groups were conducted with scientific researchers to more specifically identify attributes of a research environment. The focus group discussions uncovered a number of unique tensions in managing research, which demonstrated not only that competing values exist in the research environment, but that these values differ depending on the research objectives and tasks. For example, a tension mentioned often in these discussions was between researchers' desire and need for autonomy and management's desire and need to focus research and meet deadlines. In their study of productive climates for scientists, Pelz and Andrews (1976), found similar tensions such as these prevalent and important to consider in managing scientists because of their significant impact on performance. The findings of the exploratory study and subsequent surveys suggest that the diversity of research projects, programs and organizations can be sorted according to two primary dimensions for research objectives and two primary dimensions for describing the nature of the research work or task.

The Research Objectives Dimension

We define the research objectives dimension as Degree of Radicalness in the Scientific or Technological Advance on a continuum from incremental to radical. In slight contrast to the Schumpeterian notion of the degree of radicalness, which focuses primarily on the competitive impact of an invention (Dahlin & Behrens, 2005), we would argue that the degree of radicalness in research includes the degree of change in the state of the art, the centrality of the research problem, and the discovery of a pattern that upsets existing theory or a technology that creates a market niche. For scientific research, the task environment is the knowledge world or "the state of the art," that is, how much is known, and what is considered to be an important scientific concern or requirement.

Radical advances in science sometimes occur when a central problem is solved, such as the identification of the structure of DNA (Judson, 1979). Sometimes this also happens when a major discovery is made, such as the observation of the first candidate black hole in 1971 (Cygnus X-1), or when a research finding challenges an existing theory, such as the discovery of skeletal remains in the Americas estimated to over 1,000 years older than it was theorized that the Americas were colonized or the discovery in China of fossils of dinosaurs with feathers.

The second dimension of research objectives is the Scope of Focus, a continuum from narrow to broad, defined by the number of variables or processes or components or the number of levels or systems involved, or the extremeness of the environments of the work. In product development, this includes the number of performances affected as well as the need to change supply chains and distribution chains and create idea innovation networks or strategic alliances, as discussed in great detail by Shenhar (2001) and Hobday (1998). The challenge is adapting these concepts of scope from the product development and industrial innovation literature for scientific and technological research.

As we have already observed, the question of the amount of the scientific advance can involve multiple outcomes, that is, the number of variables or processes that are being researched at the same time. Some disciplines have a systemic quality, that is, a large number of variables have to be considered at the same time. Further, not all scientific problems can be approached with small research teams; some of them require a large scale focus. While progress in mapping and sequencing the human genome could have proceeded with small research teams, it took a large-scale inter-organizational program to coordinate a range of efforts so that the time needed to complete the entire genome was appreciably lessened. Indeed, the systemic quality of some types of research is frequently overlooked as a critical dimension to the scientific problem. One example of what might be called a systemic problem is research on the weather, which encompasses both oceanic and atmospheric systems. NOAA was created in 1970 to unify and coordinate the government's research efforts on various aspects of the global environmental system, including the National Weather Service (NWS). Altogether, the scope of NOAA's mandate necessitates quite expensive and specialized equipment and teams to collect the relevant data, including satellites, ships, buoys, and planes. Indeed, one could argue that the choice to conduct large scale data collection represents one form of a broad scope focus.

Just as science has a systemic quality in some areas that cannot be easily divided into small projects, there are technical systems that necessitate a large program of product development that may last many years. Leifer and colleagues (2000) observe in a series of case studies that this is a common characteristic of radical product innovation. Further, the research involves not simply the program, but also supporting technologies as well. For example, research on high speed trains, hydrogen-fueled cars, or fusion research also necessitates accompanying research on infrastructure, distribution and delivery systems.

Taken together, these two dimensions of research objectives, or intended outcomes, can be cross-classified to generate a typology of research objectives (see Table 1). In this manner, the choice of the relative emphasis on the radicalness of the scientific discovery or technological advance and *focus* scope generates four distinctive kinds of strategic choices. In general, the research of Shenhar (1993; 2001) on engineering projects is suggestive of how these two dimensions of research objectives can be operationalized in scientific research, where both the idea of scientific or technological uncertainty and systemic scope are in effect. In science and technology, it is the combination of objectives such as superconducting at room temperatures that reflects a radical advance.

Table 1. A typology of research objectives

	Degree of Radicalness in Scientific or Technological Advance	
Scope of Focus	**Incremental**	**Radical**
Narrow	Minor advances in a limited area, a few performances or components	Major advances in a limited area, a few performances or components
Broad	Minor advances in a system or on multiple performances of a system; Changes outside of system.	Major advances in a system or on multiple performances of a system; Changes outside of system.

The Research Tasks Dimensions

Research tasks can also be characterized by two dimensions. The first, and most obvious, dimension is the relative size of the research project, measured by the number of researchers, the number of different instruments, the number of technicians that are involved, and the number of teams and organizations involved. Furthermore, as is obvious, as the costs increase--the movement from 10 million to 100 million to one billon—there is a need for significant changes in the organization and management of the research. These changes in organization roughly parallel the movement from research project to research program or research organization, such as a national laboratory, to mission agency or inter-organizational network, as in a research consortium. At the extreme is the cost of a space shuttle, with a price tag of over 20 billion dollars expended over ten to twenty years (Gelès, 1999).

Significant work can be accomplished in small projects, such as the theoretical advances involving DNA (the double helix structure) and RNA (the messenger), which were radical advances in scientific knowledge, involved relatively small and complex research teams working within a circumscribed knowledge community (see Judson, 1979). In contrast, the mapping of the human genome necessitated a large research program that required the coordination of multiple research teams and inter-organizational relationships since there were over 3,000 genes. In systemic research, it is the large number of variables or properties of the system that require large scale programs to test and develop theories. In large organizations, programs of research are frequently housed in separate divisions because they focus on a specific area of research.

This dimension of size is similar to the fundamental distinction in the Competing Values Framework between flexibility and controlled structures, with the idea that larger projects are those that are more likely to be controlled. But the dimension of size also encompasses a number of consequences that create tensions about coordination and control mechanisms that inevitably impact on project autonomy that are discussed at greater length below. Some would classify the cost of a project as an indicator of radicalness of the innovation (McDermott & O'Connor, 2002), but we would argue this should be kept quite separate from measures of revolutionary breakthroughs in science or in technology because the cost of these projects is variable. Beyond this, the issue of the cost leads naturally into the more critical question of the size of the research project.

The other dimension of research projects is the complexity or diversity of research on a continuum from specialized to very diverse, as represented by the variety of scientific and engineering disciplines involved. This dimension is equally well established in the literatures on organizational innovation (Hage, 1999) and contingency theory more generally (Lawrence & Lorsch, 1967), as well as in science more generally (Mote, 2005). In our research at Sandia National Laboratories, we have found that many research projects had only six to ten people but in some cases, this represented six or more departments; in other instances, it reflected only one or two departments (Jordan, Hage, Mote, & Hepler, 2005). Thus, even in relatively small research projects, there can be a considerable variation in the degree of diversity in the composition of the researchers and technicians. Above, we have stressed the importance of the technicians and the equipment as one aspect of the size of the project. These factors also represent an element of complexity that is frequently ignored. For example, one of the more interesting aspects of complex research projects is the variety of research equipment that is utilized. In Latour and Woolgar's (1979) well-known study of the research laboratory, they found ten different instruments used for the purposes of measurement.

Diversity or complexity in the number of researchers and technicians can change across time. Furthermore, as research findings develop, one begins to recognize the need for still other kinds of equipment or of other kinds of expertise, that is, knowledge areas. The changes in the knowledge composition of the research project or fluidity over time are yet another measure of complexity. Again, as our research on Sandia National Laboratories research projects and programs demonstrates, typically each year, new scientific and engineering specialties were added and in some cases, others were dropped. In other words, complexity has not only a static dimension but a temporal one as well.

This dimension of structure typically measures whether expertise exists within an organization or within external organizations. Since complexity is essentially a measure of the knowledge pool of the research effort, it is to be expected that not all of the necessary skills and attributes needed are to be found in the same research unit or even within the same research organization, even in large organizations such as the national laboratories or mission agencies. The literature has consistently demonstrated that as the complexity of the research effort increases, there is frequently the need for expertise outside the research organization (Alter & Hage, 1993). In this manner, the search for additional expertise or knowledge fosters a greater external focus within the research project or program.

The two dimensions of size and complexity (see Table 2) yield four distinctive types of research projects: (1) small complex research projects; (2) large complex research programs or research organizations; (3) large specialized research programs or research organizations; and (4) small specialized research projects. Earlier, we emphasized the importance of adding the external, inter-organizational dimension to the variations on the organic model. Small complex research projects are usually connected to knowledge or practice communities, but typically in a more informal manner, and perhaps one inter-organizational relationship. In contrast, large complex research programs are more likely to be connected to set of inter-organizational relationships and maybe even a consortium, as in global alliances (Gomes-Casseres, 1996) or research consortia such as SEMATECH (Browning, 1995). Again, these are general dimensions with a considerable range of variability.

The Amount of Choice Involved in the Selection of Research Objectives and Tasks

Although we have employed the word "choice", there still remains a question of the latitude that managers have in selecting particular research tasks. For instance, the choices may be dictated to them by the environment, whether because of the agendas of funding agencies, control exerted by the state, or certain socialization practices that create a distinctive world view (DiMaggio & Powell, 1983). An example is that the culture surrounding the peer panel review process in general creates a bias toward what is often termed "normal science," that is, toward incremental advances in knowledge (Braun, 1998).

Table 2. A typology of research tasks

	Degree of Complexity or Diversity	
Size of Research Effort	**Some**	**A Lot**
Small	Specialized research projects	Complex research projects; knowledge communities
Large	Specialized research programs or research organizations	Complex research program or research organization; inter-organizational relationships

Crises can also affect the choice of research objectives and tasks. For example, the cyclical nature of environmental concerns has been manifested in rising and falling pressures on public research laboratories and private companies and shifting choices of research objectives and tasks from incremental ones, such as minor improvements in gas mileage, to more radical ones, such as the hydrogen fuel car. As this illustrates, not all pressures are necessarily directed towards radical advances in knowledge. For instance, the pressing need for immediate improvements in national security after the events of September 11, 2001 in the United States might necessitate the need to place greater emphasis on incremental advances, that is, to quickly transfer current technology into security applications (Trajtenberg, 2006).

But the most interesting constraints on choice emerge from the nature of the research problem. Natural systems that are at the extremes of scale, from the very small (subatomic) to the very large (outer space), require quite expensive equipment, numerous technical personnel and many researchers to study them. It is the case that the research problem might be capable of compartmentalization so that it can be pursued as smaller research projects. Obviously, when this is possible, the pursuit of a smaller research project presents itself as a viable research choice. But if one wants to study the system in its entirety, then either a large research organization, mission agency or a research consortium becomes necessary.

Dilemmas in the Choice of Research Objectives and Research Tasks

The influence of the external world is not the only pressure that affects the choice of the research objectives and tasks. Other, perhaps more critical, pressures are the different kinds of management dilemmas that accompany each choice. These need to be discussed in some detail because they reflect important management challenges and provide examples of how management style could make a difference.

Two very common terms in the management literature, especially in discussions of innovation, are the terms “risk” and “uncertainty”. Both of these apply to the choice of the research objectives and reflect dilemmas, particularly in the notion of how much risk one should absorb and how much uncertainty. While the terms are frequently used interchangeably in the literature, we would propose making the following distinctions: risk is a measure of the degree of radicalness in the research objective, while uncertainty is a measure of the scope of the focus on possible outcomes. A review of the literature on the management of innovation indicates that when one shifts to the language of uncertainty, the issue becomes simply a matter of counting the number of uncertainties. In other words, there are not only technological uncertainties but market uncertainties and others, although the role of the market in many areas of basic and applied research are not readily clear (Clarke, 2002). And while it follows that a large number of uncertainties, almost by definition, translates to high risk, the converse is not necessarily true. Indeed, there might only be a few technological uncertainties but still the choice to pursue a radical outcome would entail high risk. For example, Shenhar (1993) and Hobday (1998) suggest that technological uncertainty is related to the degree in which new devices, knowledge or techniques are embodied in a product. But it could be the case that one seeks a radical research outcome, yet still utilize existing technology to do so.

In short, we propose the idea of using risk and uncertainty in somewhat different ways, while still recognizing that the combination of many uncertainties with high radicalness presents the most difficult management tasks. In this manner, we are suggesting that each dilemma flows from a different set of issues than that found in new product development and industrial organization. Attempting to make a significant advance in science or technology development obviously carries a large risk of failure. Uncertainty, however, flows more from the scope of the focus that one is desiring to achieve because as the implicit number of unknowns increases, it becomes more and more uncertain as to which set is the most critical. Stated another way, when the focus is on understanding an entire system, either large or small, especially at multiple levels, there are a large number of potential unknowns or avenues of research.

One might assume that the simple solution to reducing risk is by pursuing a greater number of smaller, incremental research projects. In one sense, this is a familiar dilemma faced by managers of research in risk-averse, budget-conscious organizational environments. But there are times when progress can only be made by taking a radical approach, and “failures” can still result in a scientific advance in the sense that learning took place. In this regard, one is reminded of Thomas Edison’s remark the he did not fail, he simply found ways that did not work. Similarly, one might assume that the simple solution for reducing uncertainty is to pursue a small-scale project that is highly focused on a smaller range of unknown factors, but this can result in losing what many might call the “big picture”, a problem in any systemic science or multiple component technology.

Managerial dilemmas are not restricted simply to the choice of research objectives but also exist in the selection of research tasks. Research projects of a large size require a significant investment in management control and coordination. Such projects also typically need a substantial support staff and support systems for required services, such as accounting, human resources, and library resources. Further, larger teams also need effective leaders who can allocate resources and maintain communication and focus among team members. Finally, management also must define and communicate clear goals and strategies in order to align large groups with strategies. Indeed, the success of a large research project often depends on management correctly positioning the research to fulfill a need or fill a niche. Overall, a unifying system-wide scope makes it possible to set specific goals and track progress. The dilemma is that in increasing size and, correspondingly, internal coordination and control, there is less and less research autonomy and more bureaucracy.

Clearly, the distinction between autonomy and internal control is one of the basic structural dilemmas. Therefore, given the tight connection between size and managerial control, we have included the degree of coordination of the research project or program on the vertical dimension of the typology of structure. The desired independence and autonomy of academics are well known, as is the fact that researchers are motivated as much by the recognition of their work and the intrinsic pleasure of doing their research as by extrinsic rewards. The tension between autonomy and control also manifests itself in another way, which is the need to search for expertise as complexity increases. In so far as one locates this expertise outside the organization, another structural dilemma emerges because it conflicts with organizational autonomy given that some external control comes with the external expertise. In this regard, coordination difficulties across organizational boundaries are a dominant theme in the inter-organizational literature.

We note that there is a certain irony in this dilemma. Those projects that are already more complex because of their revolutionary strategy are precisely the ones that are most likely to recognize the need for other pools of knowledge. This flows from their aspirations and thus makes the strategic choice so critical, as Zammuto and O'Connor (1992) argued in explaining the adoption of the radical process technology of flexible manufacturing. Meeus and Faber (2006) also observe this is true on the inter-organizational side of the structure as well.

In summary, we have identified a number of dilemmas attached to both the choice of research outcome and the choice of research task within the constraints of how the context—economic, political and scientific—constrain the choices that are made. And as we combine the choice of outcome with the choice of task, these dilemmas multiply, a subject we turn to in the next section.

Constructing the Research Profiles and Management Styles

A central idea in contingency theory is that structure must follow from strategy, thus the two basic dimensions of research objectives discussed above must dictate how the research tasks should be structured. The creation of general variables that is accomplished with the typologies describing both the research objectives and the research tasks allows one to develop several hypotheses about the match between task and objective or intended outcome, the equivalent at the project or program level to the match between structure and strategy. The heart of our perspective is connecting the typology of research objectives with the typology of research tasks, which results in four kinds of research profiles. And since each choice of outcome and of task also has a set of related dilemmas, it is also the case that there are four general areas of managerial problems.

Connecting Research Tasks to Research Objectives

As we stated above, the matching of research task with research outcome is the heart of the theory of management styles. The construction of the new perspective is based on a series of hypotheses that emphasize the range of each of the dimensions used to identify research objectives and research tasks. Of these hypotheses, we present two, both of which draw heavily on contingency theory.

1. The greater the emphasis on the radicalness of the outcome in science or technology, the greater the need to emphasize the complexity or diversity of expertise in the research project, program or inter-organization network.

The more that managers choose to pursue a radical breakthrough in science, then the research project itself must become complex, particularly in terms of different types of knowledge. As discussed before, this flows from the need to synthesize different perspectives as a way of achieving the radical breakthrough. In this vein, Shenhar's (2001) study of engineering projects provides at least some evidence for the fit between the choice of a revolutionary strategy and the need for expertise; projects coded as most uncertain had the highest proportion of academic degrees.

Research findings on the positive relationship between complexity and industrial innovation defined as new products or services have been accumulated over a forty-year time period. Damanpour (1991) summarized the first thirty years of work in a meta-analysis that explored the relative importance of complexity vis-à-vis other variables while controlling for a number of alternative possibilities. Hage (1999) updated this study and noticed the absence of research on complexity and radical innovation in science. For instance, in discussions about the importance of the discovery of the double helix structure of DNA, it is often overlooked that Watson had been trained as a biologist and Crick as a physicist (Watson, 2001). In a study of the Institut Pasteur, the relationship between complexity and radical innovation holds quite strongly at the level of project (Hage & Mote, 2008).

2. The greater the emphasis on a broad scope of focus, the greater the need to increase the size of the research project, program or inter-organizational network.

As the scope of research broadens, the specific expertise that is needed may not reside within the research organization and must be sought outside the organization, either informally, as in knowledge or practice communities, or formally, as in inter-organizational alliances or consortia. The search for expertise suggests that the research unit increases in size as a consequence, moving along the continuum from project to program to inter-organizational network. The movement along this continuum is driven by the notion that scope is a measure of the number of variables involved in the research and the choice to study multiple levels, components of systems or entire systems. Many of the national public laboratories in the U.S. were created to work on difficult and intractable problems. For example, the origins of both Los Alamos and Sandia National Laboratories reside in the Manhattan Project and the drive to construct the first atomic weapon. In response, the Europeans created CERN in 1954 to reverse the brain drain of physicists to the United States and to restore and rebuild European research in high-energy physics.

In general, these types of large national laboratories and mission agencies are quite different from the smaller academic research projects funded by the National Science Foundation (NSF) or National Institutes of Health (NIH). Furthermore, within these national laboratories and mission agencies, additional distinctions can be made between problems that can be approached with small projects and those that require a more extensive program of research. If the project focuses on a system, then almost by definition, one needs a large-scale research program to handle all the aspects of the system, if not a mission agency or national laboratory. Oceanographic ships, space shuttles, and radio astronomy observatories are all examples of quite expensive equipment needed to study particular natural systems, which in turn means quite large support staff and a large number of technicians for their operation.

Another aspect of scientific research is what might be called the difficulty of the research problem, that is, studying phenomena under extreme conditions. For example, consider the case of conducting research on flora and fauna at the bottom of the ocean or the difficulty of studying black holes at the other end of the universe. Again, the presence of extreme conditions often necessitates a large scale program of research.

With these two hypotheses, we can delineate four distinctive kinds of Research Profiles, each with their own distinctive management challenges and dilemmas. We now turn to a discussion of these four Research Profiles.

FUTURE RESEARCH DIRECTIONS: RESEARCH PROFILES AND MANAGEMENT STYLES

Combining the two dimensions generates four ideal-type Research Profiles, each of which can be used to define a particular management style. In addition, because we have suggested that there were a series of dilemmas associated with the choices of the degree of radicalness and the degree of scope and similarly with the degree of complexity and the size of the project, we want to explore ways in which these dilemmas can be managed.

In Table 3, we have combined the dilemmas associated with the strategic structural choices on the assumption that structure follows strategy as we have hypothesized above. Within each quadrant are potential solutions to the dilemmas associated with each Research Profile. It is critical that one interpret this figure as suggesting dilemmas in all four quadrants. While it is easy to argue that high risk or many uncertainties pose managerial problems that necessitate solutions, it is much harder to recognize that too little risk or too few uncertainties also pose managerial problems that require solutions. In particular, the combination of little risk and few uncertainties, which is part of the Small Incremental Research Profile, raises issues about the loss of competitiveness of the research organization in which the project of this nature are located, as well as lack of understanding even by research sponsors, of the importance of some of this work.

Two important general qualifications need to be made about the four Research Profiles. First, the dimensions on which the Research Profiles are based have considerable range. In other words, one must remember that the underlying dimensions are the degree of radicalness, the broadness of the focus, the degree of complexity, and the size of the project, program, *or* inter-organizational network. Of course, this means that the specific managerial problem discussed occurs in varying quantities accordingly. Second, we recognize that each of these dilemmas exists to a certain degree in each quadrant. For example, if we focus on the Small Radical Research Profile, the two primary managerial problems are to encourage risk-taking and integration of diverse perspectives. But these same issues exist in the Large Radical Research Profile, only now they are overshadowed by the larger number of uncertainties and the greater amount of coordination needed to conduct research. In contrast, in the Small Incremental Research Profile, the motivational problems are more centered on the issue of keeping abreast of the discipline or the competition. As a consequence, a very different set of managerial solutions are needed because of the greater research autonomy. In the Large Incremental Research Profile, the same motivational problems exist but now they emerge at the level of research teams within the program and thus necessitate a different set of managerial solutions. In this respect, we associate dilemmas and Research Profiles where the problems tend to be accentuated.

To help set the stage for future research on this perspective, we identify one managerial problem associated with the choice of the research outcome and one problem associated with the choice of the research task. Many of these examples arose from the discussions in the focus groups that were conducted in the development of the research environment survey instrument discussed earlier (Jordan & Streit, 2003; Jordan, Streit, & Binkley, 2003; Jordan, Streit, & Matiasek, 2003)

Table 3. Research profiles: Strategic and structural dilemmas and associated management challenges

Strategy		*Lower Risk/Payoff*	*Higher Risk/Payoff*
	Structure	*Less Integration*	*More Integration*
Lower Uncertainty	***Less Coordination***	Small specialized projects; Value the individual; Provide professional development.	Small complex projects Encourage exploration and risk taking; Integrate ideas externally and internally.
Higher Uncertainty	***More Coordination***	Large specialized programs; Good internal allocation of resources; Good technical management.	Large complex programs; Clearly defined vision, project goals; Build strategic relationships.

Small Radical Research Profile

It goes without saying that most, if not all, researchers would like to have a radical breakthrough in their scientific or technological research, and many often tend to think that their research is geared toward that end. However, as we have suggested, it is important to view this along a dimension with considerable range. In our work at Sandia National Laboratories, we identified several examples of small projects (under 10 million funding over five years) with complex research teams that achieved radical breakthroughs, such as the project that developed a particulate trap for diesel engines that reduced particulates by 400 percent as well a project focused on the development of semi-autonomous, modular robotic control technology that greatly decreases the necessity of human control. Within NOAA, we identified a project focused on new compression methods for satellite data which achieved compression rates of almost a factor of 3 over a current compression standard (Mote, Jordan, & Hage, 2007).

Two managerial problems associated with this Research Profile are encouraging researchers to take intellectual risks and to integrate them as a team. In other words, the motivation issue is how to encourage a team of researchers to think and act boldly. The managerial solution associated with the choice of pursuing radical objectives is to create enough time to think and to explore, encourage the willingness to take risks and also have an atmosphere of challenge. A potential solution is to provide sufficient flexibility, with the resources and freedom to pursue new ideas. This suggests that it is necessary to give researchers in this profile enough time while resisting the pressure of deadlines that may be coming from those higher in the hierarchy.

As we have discussed, increasing the radicalness of the outcome requires that one increase the complexity of the research team, adding multiple disciplines, specialties within them, and technicians and their specialized equipment. But here we face a critical managerial problem associated with this Research Profile: How does one integrate across the various disciplines and equipment so that there is access to tacit knowledge? Furthermore, as Nooteboom (1999, 2000) has observed, although one creates radical innovation by increasing cognitive distance, the tendency is for people to communicate less as the cognitive distance increases. Clearly, the managerial solution is to develop a number of mechanisms to encourage the sharing of tacit knowledge (Judson, 1979), that is to encourage integration across diverse perspectives. Furthermore within the context of science and smaller research projects that address only a component of a larger system or area, we would suggest that this requires more than the classical mechanisms suggested by Lawrence and Lorsch (1967), because this involves ensuring the cross-fertilization of ideas and managing the external collaborations with various knowledge communities.

Large Radical Research Style

As one moves from a project that costs less than ten million dollars over a five year period to a program that costs more ten million dollars, and even upwards of a billion as in the example of NASA's space program, a new set of problems emerge that necessitate a different managerial style. Indeed, the increase in scale magnifies the problems described above, but alters them in interesting ways.

One of the earliest examples of a large-scale program of research is the famous Manhattan Project that produced the first atomic weapon. Since that time, large-scale research programs have been initiated to pursue a number of high-profile topics, with recent examples being the Human Genome Project and the National Nanotechnology Initiative (NNI). On a somewhat smaller scale, Sandia National Laboratories has initiated a handful of relatively large projects under the category of Grand Challenges, which are designed to be first to solve a major technical challenge. One recent example has been a research project on the feasibility of wide-scale, mainstream use of light emitting diodes (LED), which requires advances in both semiconductor design and manufacturing.

In this Research Profile, not only does one want to encourage taking risks and to integrate diverse perspectives, but one also has to be concerned about handling the number of uncertainties and the multiple research teams, including teams that cross organizational boundaries. What are the solutions? To maintain an emphasis on a radical breakthrough in a larger scale program of research, one must put forth a clear and consistent vision that provides much of the motivation for the research teams. Part of this vision is the investing in future opportunities that may provide the breakthrough in either the knowledge or technology needed to advance. The problem of handling the many uncertainties is dealt with by developing a clear set of project goals and strategies so that the many different research teams and organizations that are cooperating together in this effort can understand their responsibilities. The clarity of the vision, such as Project Apollo's explicit goal to land a man on the moon and return him safely to Earth, is critical for the success of these large scale and complex efforts. Initially, the process of determining the right research strategy and achieving a consensus on that strategy can be both time and resource intensive. Hence, continued funding over long time periods is essential. Therefore, projects and programs in this profile often endure well beyond the shorter time horizon of smaller projects, as Leifer et al (2000) have observed.

But an increase in scope of focus also creates new difficulties with managing the complexity of multiple research teams in networks of organizations, all of which have to be integrated. And in large programs and research organizations, it is important to maintain coordination and control over the research efforts. As research autonomy is reduced in this manner, so too are the opportunities to create a radical breakthrough. One of the ways in which this reduction in autonomy impacts on the creativity of the research is the inherent conflict between creativity and maintaining progress and schedules—an important issue in large complex programs. While measuring and tracking progress is an essential part of the solution to this managerial problem, it must be done in a way that does not force researchers to focus on shorter term, incremental, measurable products, benchmarks, or milestones, or take too much time away from research.

Small Incremental Research Profile

The managerial problems shift dramatically in small scale research projects designed for incremental improvements in science and technology, which, of course, are the most common ones in academia, government and in industry. Interesting examples from our work with NOAA include the ongoing, incremental improvements in the calibration of satellite data, as well as the identification of a new type of hurricane that maintains its intensity for a longer time period. A project from our work at Sandia National Laboratories also helps to illustrate this profile. This research project seeks to understand the nature of soot formation from unsteady, "dirty" flames; previous research primarily focused on the properties of steady flames generated by simple, pure hydrocarbon fuels. As these examples demonstrate, as well as many of the research projects funded by the NIH or the NSF, the research is not necessarily aimed at radical advances but at incrementally pushing the knowledge boundaries.

Since the projects in this profile are relatively smaller, the effectiveness depends much more upon the individual researcher and, at most, a small team. Therefore, one of the managerial problems in this profile is to ensure continued professional development, particularly with regard to maintaining and improving skills. Left to their own devices, researchers often tend to focus on the same set of problems across much of their professional life. In this regard, a kind of human capital decay sets in because not enough new issues are being considered to really continue to master a particular specialty. Professional development, such as continued education, seminars, and conferences, becomes a key mechanism for increasing their human capital. Individual growth is also encouraged when there is an atmosphere of cooperation rather than competition among diverse research projects in an organization. Under these circumstances, individuals are more willing to interact and talk with others about their research, particularly scientific challenges they might be facing and need help in overcoming.

Unlike the Small Radical profile, the problem here is potentially too low a level of aspiration rather than one that is too high. Again, the managerial solution to this problem is in various ways to indicate to the individual that they are valued and their work is valued even if it does not need to take intellectual risks or a broad focus. This can be accomplished by providing respect for the individual and valuing their ideas and opinions which can be achieved by allowing autonomy in decision-making, as well as more latitude in choosing areas of research. While this may appear to be quite simple, it is, in fact, often more difficult because typically the rewards are given to those scientists and research teams that make the big breakthroughs. Therefore, part of the solution includes providing greater recognition to the contributions of researchers who have not experienced similar achievements.

Large Incremental Research Profile

The combination of less radical and broad scope of focus research is frequently found in the various national laboratories and mission agencies funded by federal governments. This profile is characterized by large-scale and functionally specialized teams. For example, NOAA maintains separate research programs on both the atmosphere and the oceans, each of which encompasses a number of integrated research projects. Further, the development and implementation of a new weather satellite typically entails a substantial program that includes teams from a number of mission agencies, such as NOAA and NASA, the military and private contractors. This type of research is focused largely on incremental advances and poses different managerial challenges than research efforts focused on radical advances, such as hydrogen fuel-based cars or unmanned aerial vehicles.

Paradoxically, the large scale of research in this profile poses two managerial problems that are also observed in small, specialized projects discussed above, that is, professional development and individual respect. Given the larger scale of research, however, the existence of these two problems in this profile reflects the need to overcome the impersonality of layers of bureaucracy and, hence, suggests different management strategies. A major difference in which the handling of motivation in this profile differs from that above is that rather than value the individual or the few individuals in the team, here the problem is proper recognition of each team. Unlike the Large Radical Profile, which also emphasizes teams and teamwork, this profile emphasizes the pursuit of incremental advances, primarily on components of an overall system and use of functionally specialized teams. It goes without saying that coordination among these teams relies not only on management that is well informed about the research and rewarding and recognizing individual merit, but also on good internal allocation of funding. Cultivating well-informed, technically adept management within the research project is always difficult, for it requires skill in administration and management of people as well as knowledge and skills in the research area. This is particularly important given that the research expertise may lie in separate departments not under the immediate control of the program manager. Indeed, it is this kind of profile that led to the creation of matrix authority structures and, as is well known, this particular structure has not always functioned well because employees find it difficult to serve two masters, discipline (or department) and project (Tushman & O'Reilly, 1999; Tushman & O'Reilly, 1996). For all of these reasons, the need for clear cut measures of technical progress for each team is essential so that progress can be measured carefully and appropriately.

As indicated in the Large Radical profile, one of the major issues for large projects is the lack of research autonomy. Shenhar (2001) provides a vivid description of how managers in this profile attempt to bureaucratize the coordination of the disparate parts of the research program or teams. But, in fact, this solution is what creates many of the problems within this research profile. In addition to the impersonal effects, bureaucratic control also has negative impacts on technical management in requiring time for administrative tasks that can detract from time spent on research. Indeed, in this situation, team leadership becomes very desirable as a solution (Hage & Mote, 2008).

CONCLUSION

In summary, our argument is that the management of different types of research, including basic, applied, product development, and quality control, necessitates an adaption of existing theories of innovation and R&D management in the current literature. The Research Profiles should help with portfolio level decisions because it proposes general categories of research objectives and tasks that can be applied across various arenas of research from basic to manufacturing research to inform strategy decisions and to judge progress and effectiveness. Thus, our perspective about the diversity of research projects, contained in the Research Profiles, has an incredible spread of potential impact. Of course, a test of this assertion is the number of insights that one obtains from the perspective and below we identify and discuss three such insights.

One of the insights is the revision of the organizational learning perspective. In this respect, we would argue that the Research Profiles represent four different kinds of learning. The strategic choice of pursuing a radical breakthrough is probably going to lead to more learning than the pursuit of an incremental strategy. This corresponds to the notion of competency destroying innovation and competency enhancing innovation, which largely looks at the impact on other organizations. The Research Profiles perspective suggests that the organization that has the radical breakthrough is learning a great deal and building its capabilities in the process. As has been indicated in our discussion of each of the four profiles, whether the organizational structure and management style match the research strategy will determine if the desired scientific and organizational learning actually takes place. In each instance, different obstacles to learning present themselves. A second insight is the revision of organizational innovation theory, particularly the emphasis on the organic model, and the integration with the inter-organizational literature. The diversity of research projects shifts the attention in organizational innovation to the study of where the research is largely accomplished, while recognizing that there is not just the organic model but four profiles of research activity.

The third insight is the discussion of inherent dilemmas and problems. Rather than perceive low risk and low uncertainty as presenting no managerial dilemmas, we have suggested that they do. And unlike traditional contingency arguments that assert that the correct fit ameliorates any managerial problems, we are suggesting quite the opposite. Even with a good fit, there are still managerial problems associated with each of the four types of Research Profiles. These problems flow from the dilemmas that have been delineated. We have suggested what some of the solutions to these dilemmas are and then indicated that these define specific management styles. This is clearly not the usual perspective in either contingency theory, or management theory in general, but we would argue it is critically important given the central importance of research for both economic competitiveness and national security.

REFERENCES

Alter, C., & Hage, J. (1993). *Organizations working together*. Newbury Park, CA: Sage.

Braun, D. (1998). The role of funding agencies in the cognitive development of science. *Research Policy*, *27*, 807–821. doi:10.1016/S0048-7333(98)00092-4

Brown, E. A. (1997). Measuring performance at the army research laboratory: The performance evaluation construct. *The Journal of Technology Transfer*, *22*(2), 21. doi:10.1007/BF02509641

Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings and future directions. *Academy of Management Review*, *20*(2), 343–378.

Browning, L. D. (1995). Building cooperation in a competitive industry: SEMATECH and the semiconductor industry. *Academy of Management Journal*, *38*(1), 113–151. doi:10.2307/256730

Burns, T. R., & Stalker, G. M. (1961). *The management of innovation*. London, UK: Tavistock.

Cameron, K. S., & Quinn, R. E. (1998). *Diagnosing and changing organizational culture: Based on the competing values framework*. Reading, MA: Addison-Wesley.

Clarke, T. E. (2002). Unique features of an R&D Work environment and research scientists and engineers. *Knowledge in society, 15*(3), 58.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical? Defining and measuring technological radicalness. *Research Policy*, *34*(5), 717–737. doi:10.1016/j.respol.2005.03.009

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal*, *34*(3), 555–590. doi:10.2307/256406

DiMaggio, P., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. doi:10.2307/2095101

Gelès, C. (1999). *Managing science: Management for R and D laboratories*.

Gomes-Casseres, B. (1996). *The alliance revolution: The new shape of business rivalry*. Cambridge, MA: Harvard University Press.

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*, 109–122.

Hage, J. (1999). Organizational innovation and organizational change. *Annual Review of Sociology*, *25*(1), 597–622. doi:10.1146/annurev.soc.25.1.597

Hage, J., & Hollingsworth, J. R. (2000). A strategy for the analysis of idea innovation networks and institutions. *Organization Studies*, *21*(5), 971–1004. doi:10.1177/0170840600215006

Hage, J., & Mote, J. (2008). Transformational organizations and institutional change: The case of the Institut Pasteur and French science. *Socioecon Rev*, *6*(2), 313–336.

Hagedoorn, J., & Duysters, G. (2002). Learning in dynamic inter-firm networks: The efficacy of multiple contacts. *Organization Studies*, *23*(4), 525–548. doi:10.1177/0170840602234002

Hobday, M. (1998). Product complexity, innovation and industrial organization. *Research Policy*, *26*(6), 689–710. doi:10.1016/S0048-7333(97)00044-9

Joly, P. B., & Mangematin, V. (1996). Profile of public laboratories, industrial partnerships and organisation of R & D: The dynamics of industrial relationships in a large research organisation. *Research Policy*, *25*, 901–922. doi:10.1016/0048-7333(96)00882-7

Jordan, G. B., Hage, J., Mote, J., & Hepler, B. (2005). Investigating differences among research projects and implications for managers. *R & D Management*, *35*(5), 501–511. doi:10.1111/j.1467-9310.2005.00407.x

Jordan, G. B., & Streit, L. D. (2003). Recognizing the competing values in science and technology organizations: implications for evaluation. In Shapira, P., & Kuhlmann, S. (Eds.), *Learning from science and technology policy evaluations*. Northhampton, MA: Edward Elgar.

Jordan, G. B., Streit, L. D., & Binkley, J. S. (2003). Assessing and improving the effectiveness of national research laboratories. *IEEE Transactions on Engineering Management*, *50*(2), 228–235. doi:10.1109/TEM.2003.810820

Jordan, G. B., Streit, L. D., & Matiasek, J. (2003). *Attributes in the research environment that foster excellent research: An annotated bibliography*. Albuquerque, NM: Sandia National Laboratories.

Judson, H. F. (1979). *The 8th day of creation: Makers of the revolution in biology*. New York, NY: Simon & Schuster.

Kim, J., & Wilemon, D. (2007). The learning organization as facilitator of complex NPD projects. *Creativity and Innovation Management*, *16*(2), 176–191. doi:10.1111/j.1467-8691.2007.00427.x

Kline, S. J., & Rosenberg, N. (1986). An overview of innovation. In Landau, R., & Rosenberg, N. (Eds.), *The positive sum strategy: Harnessing technology for economic growth* (pp. 275–306). Washington, DC: National Academy Press.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills, CA: Sage Publications.

Lawrence, P., & Lorsch, J. (1967). *Organizations and environment*. Boston, MA: Harvard Business School.

Leifer, R., McDermott, C. M., O'Connor, G. C., Peters, L. S., Rice, M., & Veryzer, R. W. (2000). *Radical innovation: How mature companies can outsmart upstarts*. Boston, MA: Harvard Business School Press.

McDermott, C. M., & O'Connor, G. C. (2002). Managing radical innovation: An overview of emergent strategy issues. *Journal of Product Innovation Management*, *19*(6), 424–438. doi:10.1016/S0737-6782(02)00174-1

Meeus, M. T. H., & Faber, J. (2006). Interorganizational relations and innovation: A review and theoretical extension. In Hage, J., & Meeus, M. T. H. (Eds.), *Innovation, science and institutional change: A research handbook* (pp. 67–87). London, UK: Oxford Press.

Menke, M. M. (1997). Managing R&D for competitive advantage. *Research Technology Management*, *40*(6), 40–42.

Mohrman, S. A., Galbraith, J. R., & Monge, P. (2004). *Network attributes impacting the generation and flow of knowledge within and from the basic science community (Tech. Rep.)*. Technical Report Submitted to the Department of Energy.

Mote, J. E. (2005). R&D ecology: Using 2-mode network analysis to explore complexity in R&D environments. *Journal of Engineering and Technology Management*, *22*(1/2), 93–111. doi:10.1016/j.jengtecman.2004.11.004

Mote, J. E., Jordan, G. B., & Hage, J. (2007). Measuring radical innovation in real time. *International Journal of Technology*. *Policy and Management*, *7*(4), 355–277.

Nooteboom, B. (1999). *Inter-firm alliances: Analysis & design*. London, UK: Routledge. doi:10.4324/9780203265277

Nooteboom, B. (2000). Institutions and forms of co-ordination in innovation systems. *Organization Studies*, *21*(5), 915–939. doi:10.1177/0170840600215004

Pelz, D. C., & Andrews, F. M. (1976). *Scientists in organizations: Productive climates for research and development*. New York, NY: Wiley.

Powell, W. W. (1998). Learning from collaboration: Knowledge and networks in the biotechnology and pharmaceutical industries. *California Management Review*, *40*(3), 228–240.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, *41*(1), 116–145. doi:10.2307/2393988

Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science*, *29*(3), 363. doi:10.1287/mnsc.29.3.363

Shenhar, A. J. (1993). From low- to high-technology project management. *R & D Management*, *23*(3), 199–214. doi:10.1111/j.1467-9310.1993.tb00823.x

Shenhar, A. J. (2001). One size does not fit all projects: Exploring classical contingency domains. *Management Science*, *47*(3), 394–414. doi:10.1287/mnsc.47.3.394.9772

Shinn, T. (2002). The triple helix and new production of knowledge: Prepackaged thinking on science and technology. *Social Studies of Science*, *32*(4), 599–614.

Trajtenberg, M. (2006). Defense R&D in the anti-terrorist era. *Defence and Peace Economics*, *17*(3), 177–199. doi:10.1080/10242690600645076

Tushman, M. L., & O'Reilly, C. A. (1999). Building ambidextrous organizations. *Health Forum Journal*, *42*(2), 20.

Tushman, M. L., & O'Reilly Iii, C. A. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review*, *38*(4), 8.

Van De Ven, A. H., & Polley, D. (1992). Learning while innovation. *Organization Science*, *3*(1), 32. doi:10.1287/orsc.3.1.92

Watson, J. (2001). *The double helix: A personal account of the discovery of the structure of DNA*. New York, NY: Touchstone.

Zammuto, R. F., & O'Connor, E. J. (1992). Gaining advanced manufacturing technologies' benefits: The roles of organization design and culture. *Academy of Management Review*, *17*(4), 701.

KEY TERMS AND DEFINITIONS

Competing Values Framework: The Competing Value Framework is a theoretical framework for assessing an organization's culture. The framework consists of two dimensions, organizational focus and decision making control, from which one can derive four distinct organizational cultural archetypes: Clan, Adhocracy, Hierarchy, and Market. (Cameron & Quinn, 1998; Quinn & Rohrbaugh, 1983)

Idea Innovation Network: The idea innovation network theory is a conceptual framework for understanding the generation of innovations across organizations. The theory posits that research is differentiated among six arenas reflecting basic, applied, product development, production, quality control and commercialization/marketing (Hage & Hollingsworth, 2000).

Innovation: Involves the creation of a new idea, method, process or device. One of the earliest contributions on organizational innovation was the concept of the organic organization (Burns

& Stalker, 1961). In recent years, a great deal of attention has focused on radical innovation (Leifer et al, 2000, McDermott & O'Connor, 2002).

Networks: A network is a pattern of relationship among actors, from individuals to organizations. Within the innovation literature, the types of networks studied have included intra-organizational networks (Mote, 2005), knowledge communities (Mohrman, Galbraith & Monge, 2004; Shinn, 2002), interorganizational networks (Alter & Hage, 1993; Powell, Koput & Smith-Doerr, 1996), and strategic alliances (Hobday, 1998; Shenhar, 2001).

R&D Organizations: Research and development (R&D) organizations are public or private organizations, or subunits of these organizations, focused on the development of scientific or product innovations (Trajtenberg, 2006). Many have argued that such organizations have unique work environments (Clarke, 2002), which require different approaches to management (Geles, 1999).

This work was previously published in Technological, Managerial and Organizational Core Competencies: Dynamic Innovation and Sustainable Development, edited by Farley Simon Nobre, David Walker and Robert J. Harris, pp. 408-427, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Section 7
Critical Issues

This section contains 13 chapters, giving a wide variety of perspectives on Industrial Engineering and its implications. Such perspectives include reading in privacy, gender, ethics, and several more. The section also discusses new ethical considerations within social constructivism and gender gaps. Within the chapters, the reader is presented with an in-depth analysis of the most current and relevant issues within this growing field of study. Crucial questions are addressed and alternatives offered, and topics discussed such as creative regions in Europe, ethos as an enabler of organizational knowledge creation, and design of manufacturing cells based on graph theory.

Chapter 83
Cultural Models and Variations

Yongjiang Shi
Institute for Manufacturing, University of Cambridge, UK

Zheng Liu
University of Cambridge, UK

ABSTRACT

This chapter provides the models and variation of culture. Specifically, the following concepts will be introduced: an understanding of the original of culture and how culture is studies in the academic field; the most influential culture theories which are adapted in business and management research. A detailed exploration on the methodology, dimensions, and implications of each model will be provided. An understanding of culture's impact on manufacturing systems which stems from traditional factory-based model to inter-firm network with the increase of collaboration in globalization is given Some understanding of culture's variation in the fast growing inter-firm collaboration with case observations are also elucidated.

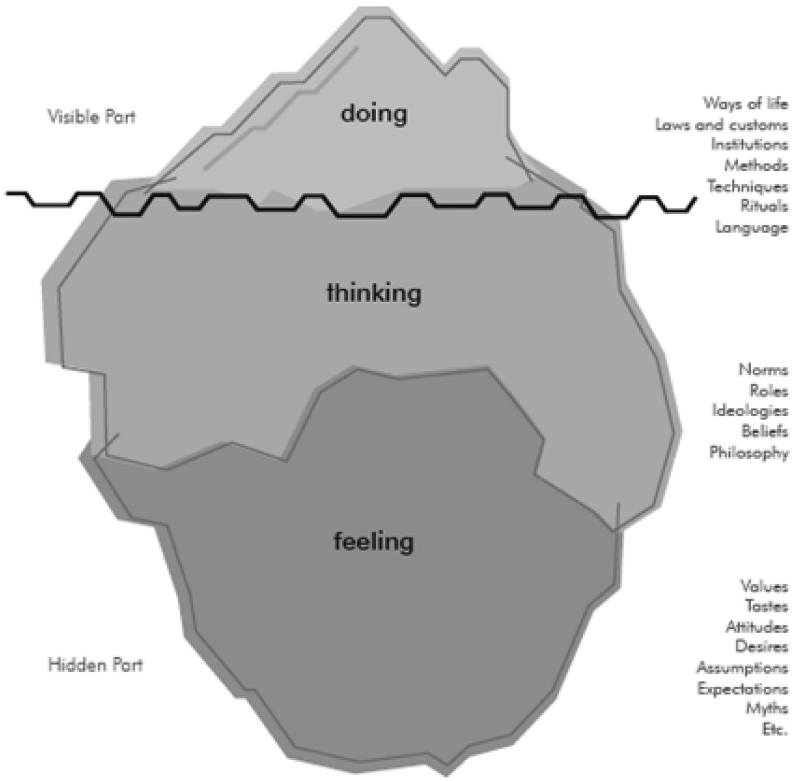
CULTURE AND ITS NATURE

Today business practices have experienced an increasing amount of collaboration as companies seek strategies to grow, restructure, and become globalized (Shi, 2005). Instead of traditional in-house production, many companies choose to collaborate as networks, particularly as international networks. The networks often include firms in different countries, and thus the interaction with companies from different national culture backgrounds becomes important. Culture influences the way companies operate. Cultural differences and conflicts become more obvious as increasingly more international outsourcing, Mergers and Acquisitions (M&As), and collaboration are developed. Culture can be defined as a system of values and norms "shared among a group of people, and when taken together constitutes a design for living" (Hill, 2000). It can be used to explain human and organizational behavior (Hofstede, 1980).

To illustrate and visualize the constructing elements of culture, several models have been developed. One of them is the "iceberg" model (French and Bell, 1979). As shown in Figure 1, it

DOI: 10.4018/978-1-4666-1945-6.ch083

Figure 1. Iceberg model of culture[1]



describes culture as an iceberg. The majority of culture, meaning to values and norms, is hidden below sea level. The top, visible area consist of behaviors, ways of life, laws and customs, institutions, techniques, rituals, language. They represent the impact of culture in daily life.

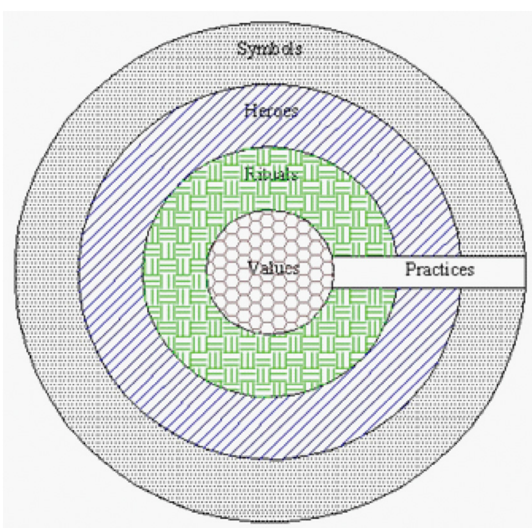
The "tree model" is another way to explain culture. As shown in Figure 2[2]. It provides the historical roots of culture—unwritten expectations, values, and norms. These factors can lead to the type of rites, stories, rules, language and behaviors, control system, and symbols.

The third model, known as "onion" (Figure 3) (Hofstede, 1994), indicates that culture, like an onion, can be peeled layer-by-layer to reveal the content. Symbols – words, pictures, or objects that carry a special meaning – are the most superficial layer. The core part of a culture is formed by the tangible symbols, heroes, story, and rituals. However, the real culture meaning of the practice is intangible. It is revealed only when the practices are interpreted (Hofstede, 1994).

In terms of cultural studies, the multi-discipline approach can be seen from perspectives of psychology, anthropology, sociology, politics, management, economy, and religion. The study of culture in psychology begins from individuals to groups and organizations, as individuals grow up within a particular culture and the accumulation of individual actions changes culture over time (Smith, 2002). Anthropology is a field of humanity starting from a group to a nation. According to anthropologists, culture is human nature. Cultural sociology is the study which uses culture analysis to interpret social life. In political science, culture is originally related to political sociology. A practical cultural approach is to identify the establishment of politics, to explain how symbols can produce political meanings, and to find out why some policies work better than others (We-

Figure 2. Tree model of culture[3]

Figure 3. Onion model of culture (Adapted from Hofstede, 1994)



deen, 2002). There are other disciplines which also explore culture factors. For example, from an economics perspective, national culture influences economic performance through institutions.

Culture as a system of values has many levels. Some typical levels are national, regional, gender, and organizational levels. These levels can overlap with each other. For example, an organizational culture can contain different regional cultures, and inside a region there can be several types of organizational cultures. In general national culture and organizational culture are most frequently studied as influential factors in business practice.

CULTURAL THEORIES AND DIMENSIONS

National culture is a frequent research in the context of international management. It is believed national level differences have a large impact on people's values as a group (Ford & Honeycutt, 1992). Many researchers try to define and measure national culture differences by establishing different sets of dimensions. Some of the most influential national culture theories are Hall's classic patterns (1959, 1969), Hofstede's cultural dimensions (1980, 1994), Hampden-Turner and Trompennars' cultural dilemmas (1993, 1997, 2000), and Schwartz value inventory (1992, 1994).

Hall's Classic Patterns

Hall is one of the earliest researchers on intercultural issues. Based on his own experience in foreign services and communication, he published two books – "The Silent Language" (1959) and "The Hidden Dimension" (1969), in which he identified two dimensions – "high vs. low context", and "polychronic vs. monochromic time orientation".

The first dimension – high vs. low context – describes the way information is transmitted. High context transactions feature pre-programmed information that is in the receiver and in the setting, with only minimal information in the transmitted message. In low context transactions, most of the information must be in the transmitted message". In this concept, countries such as the United Kingdom belong to low context culture, where information is clearly transmitted. France, on the other hand, is a high context culture, where hidden language is also used behind a conversation. Many Eastern countries also share features of high context.

Hall's second concept, "polychronic vs. monochronic time orientation", deals with the way people look at time. The monochronic time concept follows the notion of "one thing at a time", while the polychronic concept focuses on multiple tasks being handled at one time. In the culture of the UK and the USA, monochronicity is emphasized where people do tasks according to schedules, plans, and appointments based on first-come-first-serve principle. In Eastern cultures, schedules can sometimes be much more flexible. People tend to deal with emerging issues together with settled tasks. As one of the early culture studies, Hall's dimensions are very useful; however, the lack of empirical data and quantitative indications makes the concept difficult to apply.

Hofstede's Cultural Dimensions

From the 1980s, Hofstede began a survey to develop cultural models by using a large sample of IBM employees from 50 countries. Hofstede initially identified four key dimensions – "power distance (PDI)", "uncertainty avoidance (UAI)", "individualism vs. collectivism (IDV)", "masculinity vs. feminity (MAS)" – to define the differences between culture in different countries (Hofstede, 1994). Later on, the fifth dimension – "long vs. short-term orientation (LTO)" – was created in a study among students in 23 countries (Hofstede, 1994).

Power distance is defined as "the extent to which the less powerful members of institutions and organisations within a country expect and accept that power is distributed unequally" (Hofstede, 1994). It describes the hierarchical structure of a society and an organization. Uncertainty avoidance is "the extent to which the members of a culture feel threatened by uncertain or unknown situations" (Hofstede, 1994). High uncertainty avoiding countries prefer structured situations, whereas people from low uncertainty avoidance cultures are tolerant of different opinions. In individualism societies, "the ties between individuals are loose". Collectivism, on the opposite, describes a society in which "people are integrated into strong, cohesive in-groups" (Hofstede, 1994).

Hofstede defines the fourth dimension as "masculinity pertains to societies in which social gender roles are clearly distinct (Men are supposed to be assertive, tough, and focused on material success whereas women are supposed to be more modest, tender, and concerned with the quality of life); femininity pertains to societies in which social gender roles overlap (Both men and women are supposed be modest, tender, and concerned with the quality of life)." (Hofstede, 1994).

The fifth dimension was first termed as "Confucian Dynamism", and later was renamed as "Orientation" (Hofstede, 1994). Values associated with short-term orientation are thrift and perseverance. Long-term orientation is related with tradition, and fulfilling social obligations. In Hofstede's revised and expanded edition of his previous work (2010), another dimension "the nature of subjective well-being" was added to his original findings. As a societal dimension, it indicates to what degree people pursue leisure and feel happy. Hofstede's work was widely implemented in management studies. However, there is also debate that Hofstede's work is entirely based on an attitude-survey questionnaire, and it is questionable whether this is an appropriate way to find out about a nation's culture (Tayeb, 2000). The sample is also highly specific – a single industry and even a single company (Tayeb, 2000).

Trompennaars & Hampden-Turner's Cultural Dilemmas

Trompenaars and Hampden-Tuner's research is based on a ten-year period with questionnaires to 46.000 managers from 40 countries. They used a 7-dilemma model to explain culture from the perspectives of rules, individual, emotion, involvement, status, the approach to time and attitude toward the external environment. A series of seven cultural dilemmas are identified (Hampden-Turner & Trompenaars, 2000).

Universalism vs. Particularism

Universalism searches for the common similarity and standardization in a community, whereas particularism highlights the differences and uniqueness (Hampden-Turner & Trompenaars, 1993). American culture is universalism because of its "stable democracies and global cooperation – they assume that techniques can be applied universally" (Hampden-Turner, 1991). Most Catholics, Buddhist, Confucian, Hindu, and Shinto countries tend to be particularism – concerning more about personal relations (Hampden-Turner & Trompenaars, 1993).

Individualism vs. Communitarianism (Collectivism)

Individualism is associated with competition, personal growth, and fulfillment, whereas communitarianism is linked with cooperation and social concern. Most pioneer capitalist countries favor individualism, and seek personal goals and responsibilities, whereas other countries highlight team spirit (Hampden-Turner and Trompenaars, 1993).

Affective vs. Neutral Cultures

In an affective culture, people do not object to a display of emotions. In a neutral culture, people are taught that it is incorrect to show one's feelings overtly – they are aware of feelings, but keep them in control (Hampden-Turner & Trompenaars, 1993).

Specific vs. Diffuse Relationships

Cultures vary in the how specific they are. Most Protestant countries are specific. They like action, research, problem analysis, and competition. However, Catholic countries are regarded as diffuse, passionate, and romantic – they stress relationships (Hampden-Turner & Trompenaars, 1993).

Achieving vs. Ascribing Status

In a culture of achieving status, people believe there is fair atmosphere. By making efforts, one can achieve his/her goal. In ascribing culture, people are determined by their birth or origins (Hampden-Turner & Trompenaars, 1993).

Sequential vs. Synchronic

Sequential time is described as an "irreversible sequence of seconds, minutes, hours, days, months and years" (Hampden-Turner & Trompenaars, 1993). The contrast is synchronous time, which is circular or cyclical. USA people regard time as money, and race with the clock, whereas eastern people do many things simultaneously.

Inner Directed or Internalist vs. Outer Directed or Externalist

Inner direction is measured by virtues inside people – souls, wills, principles, and core beliefs. Outer direction is featured by external virtues in natural rhythms, environments and relationships (Hampden-Turner & Trompenaars, 1993). For each dilemma, according to Hampden-Turner and Trompenaars (1993), the aim is not to choose between the options, but to manage and achieve integrity.

Schwartz Value Inventory

Another theory in national culture studies was established by Schwartz (1992, 1994). Using Schwartz Value Inventory (SVT), Schwartz asked respondents to assess 57 values as to how important they felt these values were as "guiding principles of one's life". Analysis was separated into an individual-level and a culture-level. From data collected in 63 countries, with more than 60,000 individuals participating, Schwartz categorized 10 distinct values at individual-level analysis: power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity and security. Based on this structure, four value dimensions were finalized as: openness to change, self- enhancement, conservation, self- transcendence.

Seven types of national values were also found from the survey as – conservatism, intellectual autonomy, affective autonomy, hierarchy, egalitarian commitment, mastery and harmony, which were further summarised into three dimensions.

Embeddedness / Autonomy

This dimension concerns the relationship between the individual and group. Embeddedness culture emphasize group values and people's commitment to the group. In autonomy culture, people have their own uniqueness.

Hierarchy / Egalitarianism

This dimension concerns about the hierarchical structure of a society. Hierarchy represents a cultural emphasis on obeying to the unequal distribution of power. Egalitarianism emphasizes self-interests and equal opportunity.

Mastery / Harmony

Mastery refers to a culture which prefers competition and self-achievement. Harmony represents an emphasis on maintaining the world as it is, keeping a relationship rather than to change it. The SVI mentioned earlier represents a different way from previously studies in terms of methodology. It is based on a conceptualization of values and is developed with systematic measurement and analysis tools (Drogendijk & Slangen, 2006). However as a relatively new theory, the framework has not yet been tested through empirical applications enough (Drogendijk & Slangen, 2006).

Other Studies

Besides the above four key theories, there are other studies on national culture. Kluckhohn and Strodtbeck (1961), and Triandis (1972) created seven dimensions: hierarchical nature, individualism vs. collectivism, attitudes towards work, time orientation, approach to problem solving, fatalism, view of human nature. Additionally, other large scale of culture survey were taken in House et al., (1994)'s GLOBE project, and Minkov's World Values Survey (WVS) (2007). Lewis (2000) classified nations into three groups: linear-active, multi-active, and reactive, whereas Tayeb (2000) also developed seven main categories and 55 items of cultural characteristics.

Organisational Culture (Corporate Culture)

The study of organization or corporate culture has similar methodology and vision with national culture. From the foundations, culture aspects of an organization are observed through collective behavior, or the manner in which to strengthen member commitment (Durkheim, 1965). The broad studies of anthropology, sociology, political science all contribute to the literature of organizational culture. The level of analysis includes organizational level – viewing culture in its societal context, and "micro" level – concerning about individual participation and problem solving inside the organization (Fryxell, & Van Cleave, 1989).

Two general approaches of studying can be observed in organizational culture studies. The first stream views culture as "implicit in social life". The second stream views culture as "an explicit social product from social interaction as a consequence of behaviors, which contains observable forms such as customs, methods of problem solving and use of technology" (Baker, 2002). In the second approach, more practical issues are considered, such as the control power of the management groups (Martin et al., 1985), leadership, and culture change.

In terms of classifications of organizational culture patterns, Deal and Kennedy (1982) identified four key dimensions of organizational culture as: 1) values - the beliefs that lie at the heart of the corporate culture; 2) heroes - the people who embody values; 3) rites and rituals - routines of interaction that have strong symbolic qualities; 4) the culture network – the informal communication system or hidden hierarchy of power in the organisation. Denison (1990) created four types of organisational culture as: "(1) the consistency hypothesis, (2) the mission hypothesis, (3) the involvement / participation hypothesis, (4) the adaptability hypothesis". Using the four types of leaders from Greek mythology, Handy (1995) characterized organizational culture types as Zeus, Apollo, Athena, and Dionysus, and classified them into three types (Handy, 1995): 1) role cultures "Apollo"; 2) task cultures "Athena"; 3) existential design cultures "Dionysus"; 4) club culture "Zeus". Cameron and Quinn (1999) categorized organizational culture into four types – hierarchy, clan, market and adhocracy, Schein (1992) developed dimensions of organizational culture into: 1) external environments; 2) managing internal integration: 3) reality, truth, time, space, human nature, activity, relationships.

The issue of culture change is also highlighted in qualitative studies. Trice and Beyer (1993) pointed out the basic types of culture change such as the revolutionary, comprehensive, and gradual efforts. In Schein (1999)'s view, the strategies are: (1) unfreezing the old culture and creating motivation to change; (2) capitalizing on propitious moments; (3) making the change target concrete and clear; maintaining some continuity with the past; (4) creating psychological safety through a compelling positive vision, formal training, informal training of relevant groups and teams, providing coaches and positive role models, employee involvement and opportunities for input and feedback, support groups, and addressing fears and losses head on; (5) selecting, modifying, and creating appropriate cultural forms, behaviors, artifacts, and socialization tactics; (6) cultivating charismatic leaders; (7) having a realistic and solid transition plan; and, (8) exercising risk management by understanding and addressing the risks and the benefits as well as the potential inequitable distribution of these risks and benefits.

MANUFACTURING SYSTEM AND CULTURE

The linkage between culture and manufacturing system has drawn attention in both research and business practice. Culture is formed by nature forces such as environment change, and human force including war, immigration, trading, and technology (Hofstede, 1980). It influences the way people behave within a family, community, organization, region, and nation. On a system level, it also forms the way of education, family patterns, political structure, business routines, legislation, and communication within a nation and society (Hofstede, 1980). These systems will eventually influence business practice such as innovation pattern and relationship management. Figure 4 provides a framework to indicate culture's origins and its impact.

Figure 5 is the conceptual framework of manufacturing systems. Traditionally, culture influences operations in a single factory in terms of quality management and human resource management. For example, the Lean concept from Japan is much related with the Japanese culture of collectivism, and high uncertainty avoidance. The designing of shop floor layout can also reflect individual or collective culture. More studies are carried out on the impact of national culture on the using of Manufacturing Resource Planning (MRP) and Total Quality Management (TQM). As factories disperse among various countries, to some extent national culture differences can determine the coordination of new product development and production management in different locations. The issue of cross-border Mergers & Acquisitions (M&A) also includes many culture elements as host country and home country can differ in many ways. There are many studies on the relationship between the preferences of foreign investment mode, local management, innovation patterns with culture differences by using the scores provided in Hofstede's culture theory.

Figure 4. A system of culture (Compiled from Hofstede, 1980)

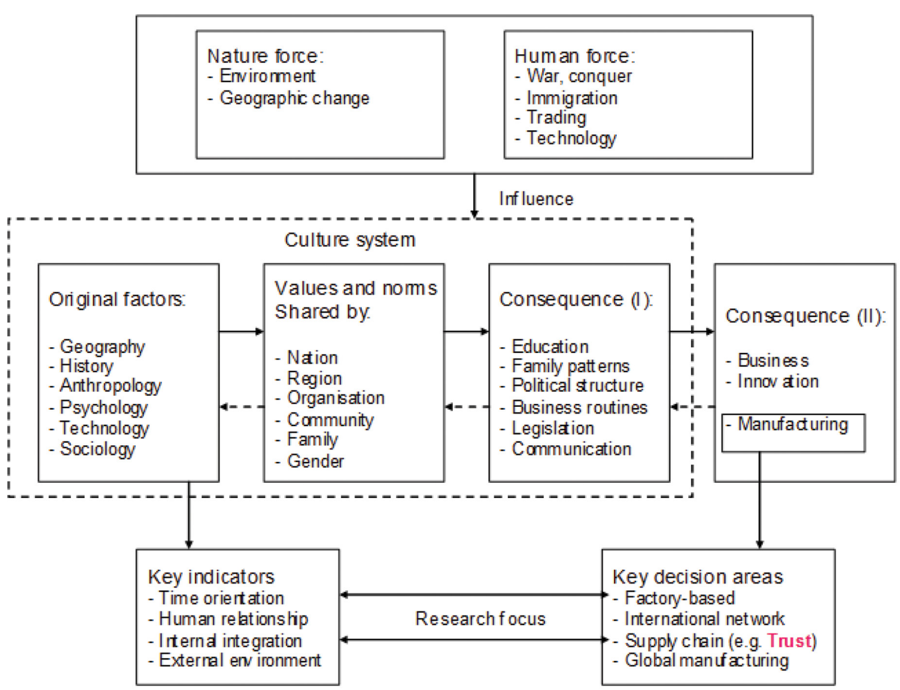
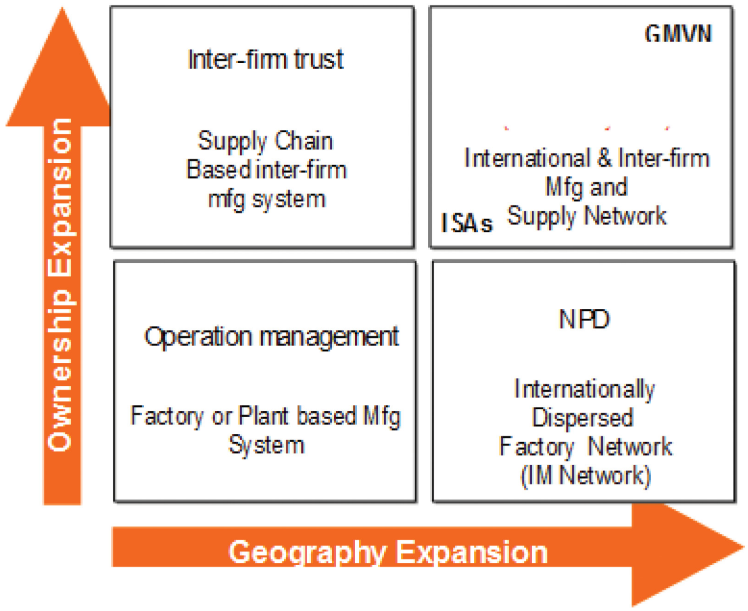


Figure 5. International manufacturing system framework. Adapted from Shi, 2005.



In the vertical axis of Figure 5, the supply chain requires for understanding and coordination of different firms who have different values and culture background. Among the inter-firm issues are trust, technology transfer, and learning. Culture differences can result in different understanding and evaluation methods of trust. For example, the inter-firm trust in western culture rely more on competence and clear contract, but in China, personal relationship is more important. To combine the inter-firm and intra-firm issues, globalization and the global manufacturing virtual network have much interaction with culture in terms of information management, network structure, coordination and control.

INTER-FIRM RELATIONSHIP AND CULTURE WITH INDUSTRY OBSERVATIONS

The impact of national culture on inter-firm relationship becomes an important issue as companies now operate within a network. There is much collaboration between firms from different national backgrounds. Understanding cultural differences can help to solve problems, and failure to do so may cause misunderstanding, conflicts, and difficulty reaching an agreement. This happens not only in traditional industry sectors, but also in new industry sectors. For example, using the creative industry in the past, the product of animations was done primarily via in-house production. Since the 1980s, there has been a trend towards outsourcing and international collaboration. In China, due to the lack of original design capability during the 1980s, most animation companies are studios which focus only on production. As a result, many western and Japanese animation designing companies choose to collaborate with Chinese studios to enter the Chinese market, and to cut down production cost. When studying into these Chinese animation companies, it is found that they are experiencing culture differences between the Chinese and western culture.

One case company from China collaborates with both Chinese and USA partners; however, the collaboration followed different processes.

One chief executive said: "In China, we first find friends; then perhaps we choose the most capable partner within that community. Sometimes we do not establish any contract at all. It is more about personal interaction." In this way, they join animation industry clusters, keeping close to other partners, and getting to know more friends through the intermediate introduction. They build up with many Chinese partners who may not be capable enough to design a full animation production. However through co-learning and sharing resources, they develop a competence together.

Meanwhile, through 10 years of international outsourcing collaboration, they also realize that "when we collaborate with western partners, we should always focus on our capability. No matter how well people know us, we must show them evidence that our technology and skills are satisfactory. There is a strict process and a standard contract to identify roles, responsibilities, technology issues, cost, IP." In order to attract more western clients, this company pays much attention to competence improvement by:

- Using the latest 3D drawing software and train employees with skills
- Developing a IP protection system in terms of hardware and software
- Exchanging staff overseas to learn cross-culture management
- Other training program such risk management
- Continuously improving knowledge and new production technology

This case shows the different approaches of the Chinese and USA way of developing inter-firm trusting relationship. Chinese companies prefer relationships first, and then co-develop a competence, whereas companies in the USA highlight the competence and standard contract first, and may develop long-term interaction later. To trace back the national culture root of Chinese and USA companies, several features can be learned through previous culture theories. A comparison can be made as in Table 1.

In general, Chinese culture has the feature of long-term orientation, polychronicity, high power distance, collectivism and diffusion. It emphasizes group goal, teamwork, and personal relationships. When developing inter-firm relationships, it tends to develop personal interaction with a long-term orientation. Additionally, tasks may not be allocated clearly. Instead, co-evolution and sharing is more desired. In contrast, the USA has the culture of short-term-orientation, mono-

Table 1. Culture differences in China and USA

Culture dimensions	China	USA
Long vs. short-term	Long-term (118) (Hofstede)	Short-term (29) (Hofstede)
Monochronic vs. polychronic	Polychronicity (Hall)	Monochronicity (Hall) Sequential (Trompenaars)
Power distance	High (80) (Hofstede) Particularism (Trompenaars)	Slightly low (40) (Hofstede) Universalism (Trompenaars)
Individualism vs. collectivism	Collectivism (20) (Hofstede)	Individualism (91) (Hofstede) Individualism (Trompenaars) Autonomy (Schwartz)
Masculinity vs. feminity	Medium (66) (Hofstede)	Medium (62) (Hofstede) Mastery (Schwartz)
Uncertainty avoidance	Low (30) (Hofstede)	Slightly Low (46) (Hofstede)
Specification vs. diffusion	Diffusion (Hall)	Specification (Hall)

chronicity, low power, individualism, masculinity and specification. It highlights the clarification, transparency of documents (in the case of contracts), a clearly defined technology, and individual responsibility. Therefore, companies in the USA may visit their partners directly to make sure that a real competence exists. It may also have a standard process of sample making, technology definition, and partner selection.

The success of this Chinese company gives a good example of handling the culture differences between the East and the West. It focuses on conflicting values and adapts itself to the new context. The company does well because it keeps the culture differences clearly in mind and respects the values from its partners. To achieve effective communication, it sends people to learn American culture in USA, and invite partners to experience Chinese culture. Other companies may still struggle with cross-culture relationship. Sometimes they turn to professional consultancy or local agency for help. A comprehensive understanding of culture theory is foundamental. Furthermore a real practical knowledge of how culture influence business behavior can be enriched through international management experience.

CONCLUSION AND FUTURE AREAS FOR CULTURAL STUDIES

This chapter first introduces the concept of culture from the perspectives of typical models, disciplines, and levels. Then particularly some most influential studies on national culture and organizational culture are provided with details of definition, dimensions, and meaning. From the literature review, it is clear that researchers explore culture mostly from aspects of national culture and organizational culture with dimensions. To finalize the similarities among the dimensions by its meanings, three types of culture can be summarised as time orientation, internal integration, and external adaptation (Table 2).

1. **Time orientation:** related with people's attitude towards time. Dimensions include the "long vs. short term orientation"; "monochronic vs. polychronic", which are combined from Hall, Hampden-turner, and Hofstede's dimensions.
2. **Internal integration:** refers to the issues related to the relationship in human organizations, such as the hierarchical structure, the emphasis on competition, and group relationships. Dimensions are the power distance, individualism vs. collectivism, and masculinity vs. feminity.
3. **External adaptation:** focuses on the interaction between human society and the outside environment. Details are uncertainty avoidance, and the content of communication (specification vs. diffusion) can also be included as external issue.

The later section of the chapter discuss how these dimensions can be adopted as national culture variables when assessing the interaction between culture and issues of business management. Particularly an industry observation was given to show the culture factor in inter-firm relationship management in the context of the USA and Chinese collaboration. The values and processes indicate different priorities of Chinese and American companies.

Following the literature, this chapter also provides preliminary thoughts on the impact of

Table 2. Dimensions of national culture

Types of culture	Key dimensions
Time orientation	Long vs. short term orientation
	Monochronic vs. polychronic
Internal integration	Power distance
	Individualism vs. collectivism
	Masculinity vs. feminity
External adaptation	Uncertainty avoidance
	Specification vs. diffusion

culture on manufacturing system, which grows from individual factory based system towards inter-firm and intra-firm international network. Some industry observations on the cross-culture inter-firm relationship issues are also explored.

There are some future research areas for culture, which are closely related with the emerging issues of business. One important phenomenon now is the emerging countries MNEs. In the past decade, many of these companies have emerged from China and India. They develop quickly, and extend resources quickly through M&A. However, some of the companies failed to conduct local people management, while others may still struggle with local adaptation. Culture differences between the host and home country contribute a lot to the problem. How to use culture dimensions to help those emerging countries MNEs to integrate resources after M&A, and build up local relationship effectively can be of much practice values in the current business climate. Also, as manufacturing nowadays aims to achieve a more sustainable system, the issue of green management, which involves the interactions with environmental and societal factors become important. Different countries may view differently towards green issues, and thus conduct different practices. There is a need to link current culture theories with sustainability dimensions. The topics of "what are the new requirements for culture" in this sustainable system will also be of great value in future research.

In terms of culture studies itself, a multi-discipline is needed. These can help to enrich the existing theories with a broader view. More qualitative research is also required to understand the process of cross-culture management in the current international business climate.

REFERENCES

Baker, K. A. (2002). *Organisational culture*. Retrieved from www.wren-network.net/resources/benchmark/11-OrganizationalCulture.pdf

Cameron, K. S., & Quinn, R.E. (1999). *Diagnosing and changing organisational culture: based on the competing values framework.* Reading, MA: Addison-Wesley Publishing Company.

Deal, T. E., & Kennedy, A. (1982). *Corporate cultures: The rites and rituals of corporate life*. Reading, MA: Addison-Wesley Publishing Co.

Dension, D. R. (1990). *Corporate culture and organizational effectiveness*. New York, NY: John Wiley & Sons.

Drogendijk, R., & Slangen, A. (2006). Hofstede, Schwartz, or managerial perceptions? The effects of different cultural distance measures on establishment mode choices by multinational enterprises. *International Business Review*, *15*(4), 361–380. doi:10.1016/j.ibusrev.2006.05.003

Durkheim, E. (1965). *The elementary forms of the religious life (J. W. Swain Trans.)*. New York, NY: Free Press.

Ford, J. B., & Honeycutt, E. D, (1992, November/December). Japanese national culture as a basis for understanding Japanese business practices. *Business Horizons*.

French, W. L., & Bell, C. H. (1979). *Organization development*. New Jersey: Prentice Hall.

Fryxell, V. C. (1989). The foundations of the organisational culture literature: an integrative Framework. *International Journal of Values Based Management*, *2*(1).

Hall, E. T. (1959). *The silent language*. New York, NY: Double Day.

Hall, E. T. (1969). *The hidden dimension. Man's use of space in public and private*. London, UK: Bodley Head.

Hampden-Turner, C., & Trompenaar, F. (1993). *The seven cultures of capitalism*. New York, NY: Currency/Doubleday.

Hampden-Turner, C., & Trompenaar, F. (1997). *Riding the waves of culture: understanding cultural diversity in business*. London, UK: Nicholas Brearley.

Hampden-Turner, C., & Trompenaar, F. (2000). *Building cross-cultural competence*. West Sussex, UK: John Wiley & Sons, Ltd.

Handy, C. (1995). *Gods of management: The changing work of organisations*. New York, NY: Oxford University Press.

Hill, C. W. L. (2000). *International business. Competing in the global marketplace*. New York, NY: McGraw-Hill.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Culture and organizations: Intercultural cooperation and its importance for survival* (3rd ed.). New York, NY: McGraw Hill.

Hofstede, G. H. (1980). *Culture's consequence: International differences in work-related*. Beverly Hills, CA: Sage Publications.

Hofstede, G. H. (1994). *Cultures and organizations: Software of the mind: Intercultural*. London, UK: HarperCollins.

House, R. J., Hangers, P., Javidan, M., Dorfman, P. W., & Gupata, V. (2004). *Culture, leadership, and organisations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications.

Kluckhohn, F., & Strodtbeck, F. L. (1961). *Variations in value orientations*. Westport, CT: Greenwood Press.

Lewis, R. D. (2000). *When cultures collide: Managing successfully across cultures*. London, UK: Nicholas Brealey.

Martin, J. Stikin, S. B., & Boehm, M. (1985). Founders and the elusiveness of a cultural legacy. In P. J. Frost, et al. (Eds.), *Organisational culture*. Beverly Hills, CA: Sage Publications.

Minkov, M. (2007). *What makes us different and similar: A new interpretation of the world values survey and other cross-cultural data*. Sofia, Bulgaria: Klasika y Stil.

Schein, E. H. (1992) *Organisational culture and leadership.* San Francisco, CA: Jossey-bass Publishers.

Schwartz, S. H. (1992). *Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Advances in Experimental Social Psychology*. San Diego, CA: Academic Press.

Schwartz, S. H. (1994). Beyond individualism/collectivism: New dimensions of values In Kim, U., Triandis, H. C., Kagitcibasi, C., Choi, S. C., & Yoon, G. (Eds.), *Individualism and collectivism: Theory application and methods*. Newbury Park, CA: Sage Publications.

Shi, Y. (2005). *Contractual manufacturing service providers in different industries: A key co-founder of the global manufacturing virtual network (GMVN).* Conference paper (the 12th International EurOMA Conference).

Smith, P. B. (2002). Levels of analysis in cross-cultural psychology. In W. J. Lonner, D. L. Dinnel, S. A. Hayes, & D. N. Sattler (Eds.), *Online readings in psychology and culture*. Retrieved from http://www.wwu.edu/~culture

Tayeb, M. H. (2000). *International business: Theories, policies and practices*. New York, NY: FT Prentice Hall.

Triandis, H. C. (1972). *The analysis of subjective culture*. New York, NY: Wiley.

Trice, H., & Beyer, J. (1993). *The cultures of work organisations*. Englewood Cliffs, NJ: Prentice Hall.

Wedeen, L. (2002, December). Conceptualising culture: Possibilities for political science. *The American Political Science Review*, *96*(4). doi:10.1017/S0003055402000400

ENDNOTES

[1] Adapted from online source: www.intercultures.gc.ca

[2] Source: http://www.army.mil/ESCC/cm/model5.htm

[3] Adapted from online source, http://www.army.mil/ESCC/cm/model5.htm

This work was previously published in Cultural Variations and Business Performance: Contemporary Globalism, edited by Bryan Christiansen, pp. 278-291, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 84
New Design Paradigm:
Shaping and Employment

Vladimir M. Sedenkov
Belarusian State University, Belarus

ABSTRACT

The multiple shortcomings of the current Design paradigm manifest the need of its modification. Our objective was to find out an appropriate mechanism. But such a mechanism could not be revealed without assistance of a Design theory. The emergent dilemma – to use one of the available theories or develop a new one – was resolved by choosing the third way: rearrangement of the material at hand on modularity principles with initiation of fundamental (systemic) Design theory module via identification of its paradigm. While doing this, we had to overcome a number of delusions ingrained in engineering design, concerned firstly with design problem, process and design representation. To push these efforts forward, a scientific base named Continuous Process Theory had been developed. Systemic module initiation enabled to define a paradigm of the second Design theory module – the sought-for Design practice paradigm. Discussion on the outcomes of this definition rounds off this chapter.

INTRODUCTION

The Way to Reveal an Adequate Design Paradigm

The title of the chapter declares generation of a new design paradigm – the set of practices that define and drive (implements and manage) design process (*DPR*). This assumes that a certain version of the paradigm is available. Indeed, it cannot be out of place because of many decades of everyday designing. Design paradigm ("applied theory") is crystallized by design practice and is needed for practitioners in each design field – whether it be mechanical engineering, information science, architecture, chemistry, nano-technology or something else. Design paradigm is the base of design methodology within a given domain and the benchmark for developers of computer-aided design facilities.

DOI: 10.4018/978-1-4666-1945-6.ch084

In Design theory absence, Design paradigm formation was spontaneous: the values of its descriptors mirror a specialized empiricism, intuition, experience, borrowings, parts of theories possessed by other disciplines, etc. This has stipulated a number of weak points of the paradigm. With the reference to design in mechanical engineering, recall some of those:

- it employs semi-intuitive design language;
- supports mainly adaptation design;
- generates ill-observable, non-holistic *DPR*, which is equally insufficient for learning and teaching and for the most part implicit;
- has no ideas how to keep the *DPR* complexity to a manageable level;
- structure synthesis problem remains unsolved;
- the role of computer in designing is obscure and insufficient;

Thus, modification or replacement (radical modification) of design paradigm is anticipated and in demand. However, there is no a regular mechanism of paradigm improvement, which should concurrently be a mechanism of paradigm identification forgoing to its change. Let us try to find out such a mechanism.

To be analyzed, evaluated and modified, the paradigm should have a sort of representation. We associate with any Design paradigm representation a set of descriptors or *paradigmants* – the certain characteristics of a paradigm, which take on one or another value. Thus, paradigmants characterize via their values this or that paradigm during a certain period of time. For instance, Design paradigm is concerned with such paradigmants as *notion base of design language* (formalized ∨ intuitive ∨ semi-intuitive), *the mode of structure synthesis problem realization* (explicit ∨ implicit), *design system architecture* (an hierarchy of subsystems ∨ another) and others.

If paradigm representation is available, we define a paradigm modification as the change of values for one or a subset of its paradigmants. To regularize the way of paradigm change, we distinguish within its representation a minimal subset of paradigmants sufficient for a unique paradigm identification – call it *design paradigm signature* (*Sg*). Then change of values for paradigmants, which are beyond the signature, would signify the paradigm modification, while the change of value for at least one signature's paradigmant replaces the paradigm. A signature considered without its paradigmants values is called a signature platform or *meta-signature* (*mSg*). Choosing different variants of *mSg* attribution, the produced signature alternatives (paradigm identifiers) could be compared and paradigm assessed as a whole.

It should be noted that the outlined mechanism of paradigm identification and modification has a heuristic base – the choice of both paradigm representation and modification rests mainly on experience, intuition and experiment. This does not make the mechanism reliable. Besides, it generates only paradigm clones according to given representation – this deprives the mechanism of practical value. Generation of paradigm versions becomes possible after changing the course of identification for an opposite one. This means that paradigm representation initially is unavailable and its deriving begins with identification of Design paradigm meta-signature and signature. Then the paradigm representation will be obtained by deployment of its signature. But such a systematic way of Design paradigm handling needs for feeding it by a resource of paradigm *mSg* generation. Such a resource could be provided by a Design theory only.

TOWARDS DESIGN THEORY PARADIGM IDENTIFICATION

Requirements to Design Theory

Does the stated need for Design theory signify development its new version or sampling from the list of currently available Design theories is also possible? To make decision, let us lay down three clue, in our view, requirements to Design theory.

1. First of all, the theory has to have a base of scientific principles. "The teaching of engineering subjects such as thermodynamics, mechanics, and materials are based on established principles like the first and second laws of thermodynamics, Newton's laws, and atomic and molecular theories of matter… Engineering design lacks sufficient scientific foundation, and without an adequate base of scientific principles, engineering design education and practice are too much guided by specialized empiricism, intuition, and experience" (Dixon, 1988, p. 145).
2. To be accepted by practice, a theory should answer its questions, solve its problems, and uppermost provide a regular mechanism of Design paradigm identification (attribution of *mSg* paradigmants) and modification. In case of design practice, its *mSg* is invariant and fixed a priori: i) constructive design definition; ii) a priori described process that provides the sought-for design, i.e. *meta design process* (Sedenkov, 2008, p. 93); iii) a system that supports meta process run or *design system*. Combining the requirements 1 and 2, we assert that the base of scientific principles should be aimed firstly at beneficial identification of Design paradigm (a proper *Sg* of Design practice paradigm).
3. The requirement of a viable theory organization is the last one in the triple. The origin of this requirement needs for comments.

We have in focus a Design theory, which should have its own paradigm – a set of ideas that define a scientific discipline during a particular period of time. This theory should identify an adequate design paradigm that is the part of itself, as the identified paradigmants of Design paradigm signature will become paradigmants of Design theory paradigm. This separation of Design theory and Design paradigm in their unity gives rise to the idea of modular structure of Design theory where paradigms of two main modules – Design theory and Design practice – define accordingly fundamental and applied aspects of a single theory.

Indeed, distinct components of Design theory may have substantially different dynamics of development. For instance, there are highly dynamic divisions aiming at knowledge handling. Next, market dynamics of computer aided design facilities and value dynamics of Design paradigm paradigmants, which are beyond the signature, interrelated closely and stimulate each other. Besides, the status of interdiscipline accepted by designing (Cross, 2002) obligates it to interact with a dynamic list of contractor disciplines. In this situation, a single Design theory has been either substantially loose or inflexible. The first variant makes a theory be of little use for design practice while the second one entails its frequent revisions and disintegration into a number of short-lived mini-theories supporting particular design stages (DfX). Between one-piece ("General") and mini-theories, there is a class of "heuristic" theories for particular design fields.

In case of modular structure of Design theory, the paradigms of modules get the autonomy in the sense of possibility to vary within the values of their current signatures. The paradigm change of Design theory itself (replacement of its signature value) occurs either on a signal "underneath" (when the paradigms of modules are not sufficient any more), or "from above" (with gathering new knowledge about signature's paradigmants possessed by Design theory paradigm and/or with alteration of scientific principles). It seems of importance to reserve the space for paradigmatics variations in each of modules as it should ensure simultaneously both the theory stability and its progress without strong conflicts between them.

Choice Testing

Let us see, whether some of the available theories could be suited for our objectives. To avoid single analysis (it is beyond the scope of this chapter), we shall manipulate the two above mentioned classes testing those for compliance with the stated requirements.

The class name "heuristic" indicates not only the origin of theories "from practice" but the domination of heuristics in their statements as well. The prevalent method of such theories creation is induction, which begins with generalization of design experience. With the terminology introduced, one may state that this class of studies (Lossack&Grabowski, 2000; Eder&Hosnedl, 2008) is characterized by the attempt to elevate applied aspect of design theory to the rank of fundamental. In other words, Design paradigm signature should become a core of Design theory paradigm signature. But as initial component is not reliable, the result proves to be artificial as well. In so doing, the weak points of design paradigm (for instance, the transfer of concepts related to representation and analysis of available products to synthesis of nonexistent products) turn to be weak points of Design theory.

The alternative class of theories ("General Design Theories") stresses the fundamental aspect putting off the handling of applied one to the end of theory formation. But when the theory comes to the focus of practice, it turns out that the applied aspect of the theory does not suit for complete and consistent use (Yoshikawa, 1981; Hatchuel, 2003). Thus, though the maxim "a theory will help advance the practice of design" (Dixon, 1988, p. 148) became a generality, it does not follows from the theory itself *how* and *by which* the advancement should come.

Having compared the properties of Design theory classes with the triple of requirements (sound scientific foundations, aimed at Design paradigm identification, and modularity), we have to conclude that the choice in the space of available Design theories could hardly be resultant to our end.

What Then, a New Theory?

However, the conclusion about necessity of new Design theory creation would be premature as well. In the light of modular structure and relative autonomy of both – modules evolution and its rate, theory shaping and its validation is the issue of time and efforts of the entire design research community. On the other hand, odd parts of a theory are in existence (Design paradigm, design methods, design knowledge, etc.) but those are either inadequate (like Design paradigm) or technologically imperfect in practical use (as knowledge or design methods).

Under this situation, it seems rational to rearrange the available material on the principles of modular theory organization. Then the first module should be established. It has doubling the parts of an attractor for odd elements of the theory and a platform for creation and synergetic operation the rest of the theory modules. This gives grounds to consider not a theory creation but its *initiation*, that is shaping a module of fundamental theory aspect or, hereinafter, a *theory platform*.

The Platform Based Design Theory Organization

The Design theory platform should provide a mode of Design paradigm signature identification and so is initially aimed at establishment and support its applied aspect. To develop the platform means firstly to propose a meta-signature of its paradigms. Then we have to initiate an adequate signature via attribution *mSg*'s paradigmants. The signature of Design theory platform will ensure the Design paradigm signature identification. The latter also has *mSg*, which is known a priori and was mentioned above.

We complete to plan a modular structure of Design theory by introduction of its third module.

It will be a meta knowledge base that involves design methods, design knowledge, computer software, contractor disciplines, multidiscipline models and will be provided with control facilities to support knowledge access, knowledge storage and knowledge maintenance. We call this module "Design knowledge management system". Its meta-signature consists of three paradigmants: knowledge structuring, knowledge storage and knowledge access.

Thus, the only way to obtain Design paradigm signature (attribution of *mSg* paradigmants) is the initiation of constructive version of Design theory via identification of paradigm signature for the main module of the theory. In turn, Design paradigm signature has been a trigger for shaping the third module of the theory via identification its signature. Such a "modular-trigger" concept of theory development removes from agenda the necessity of General Design Theory. The latter can be changed for the new entry – a unified Design theory platform. On the other hand, it has been doubtful to have internal Design theory for each particular design field – with the use of master module they can adjust two other modules from the triplet and get a required version of Design paradigm for the domain.

The reminder of this chapter is organized as follows. In the next part we briefly review the proposed science base for design – Continuous Process Theory and the mode of problems tackling when those have the same insolubility (of the second kind) as a design problem. The third part addresses to establishment of Design Theory platform that paves the way to adequate design paradigm identification. The proper design practice paradigm revealing is the content of the forth part, while the main outcomes are discussed in the fifth part.

TOWARDS DESIGN THEORY PLATFORM: THE BASE OF SCIENTIFIC PRINCIPLES

Continuous Process Theory

Dixon (1988) states that "design is a process, and processes are not the usual subject of theoretical formulations" (p. 146). A process is a continuous entity and therefore it should have a continuous representation. Discrete representation of a process – by an hierarchy of sub-processes, for instance – mostly inefficient. *Continuous Process Theory* (*CPT*) is just intended to serve for the part of adequate scientific principle for design theory development. Here are two of the most important applications: (1) with the help of *CPT*, we shall show a design problem membership of the class of problems with insolubility of the second kind and outline the way to get an answer to such problems; (2) structure synthesis problem realization needs for continuous representation of a design, and we shall get it using (continuous) representation of *operation process* of the required product.

The starting point of *CPT* is representation of any process *PR* by its scheme:

$$PR=(D, P), \tag{1}$$

where *P* stands for a *processor* that performs transformation of energy, raw materials, information or products entering its input (I^P), and *D* stands for a *procedure* that describes the function of *P* over its I^P. *D* and *P* are referred to as a process *object* and *subject* respectively. Thus, the *subject matter* of *CPT* is a representation of complex process in a continuous mode. The *CPT method* – the scheme technique of process – is characterized by the following:

- A set of process schemes is added with a number of binary relations;
- Process schemes linked by distinguished relations make up a *structure*;

- The rules for structure shaping and conditions for its operability are stated.

The *CPT goal* is a constructive proof for a process (induced, for instance, by a declared problem and represented by the scheme) its runability through building up upon its scheme an operable continuous structure of processes.

There are two relations appropriate for making up the structures: providing relation or **p**-*relation* and relation of determination or **d**-*relation*. PR_1 and PR_2 are linked with **p**-relation $(PR_2 \xrightarrow{p} PR_1)$ if the output of PR_2 serves for the input of PR_1. If the output of PR_2 becomes a scheme component of PR_1 (*D* or *P*), these two processes are linked by **d**-relation $(PR_2 \xrightarrow{d} PR_1)$.

A set of processes (or their schemes) continuously linked by **d**- or **p**-relation forms an *elementary structure* of processes (or process schemes). This structure is represented by a plain graph, the nodes of which serves for the processes and each arc is a cross-linking relation. Elementary structures have an *order n*, equal to *1* (Figure1).

Each processes structure has a *core* (PR_1), a *tail* (PR_n) or *multitail* (PR_2, PR_3) and a *body* (the part between the core and tail). Elementary structures generated by one of the relations can form a new structure by the alternative (orthogonal) relation. The new structure gets $n=2$ and is represented by a *super-tree* (*S*-tree) – an arc bichromatic (2D) tree where each node is an ordinary graph (Figure 2*a*). Relations between members within such non-elementary structure of processes are relations between the members' cores.

The structures with $n=2$ may serve for the members in a structure of the next order. A usual motive for the third order structure generation is the synthesizing a required process from scratch. To this end, let us associate with each process scheme a level of its *uncertainty* (*UL*) as *UL* of the scheme's components.

Figure 1. First order process structures on **d**- *and* **p**-*relations*

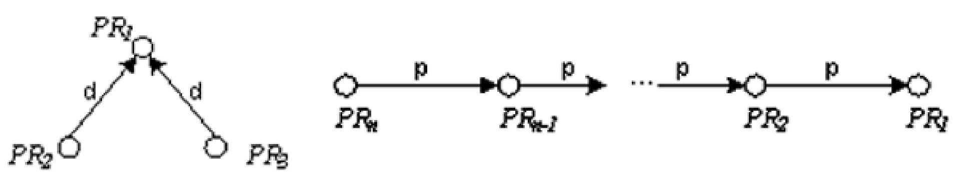
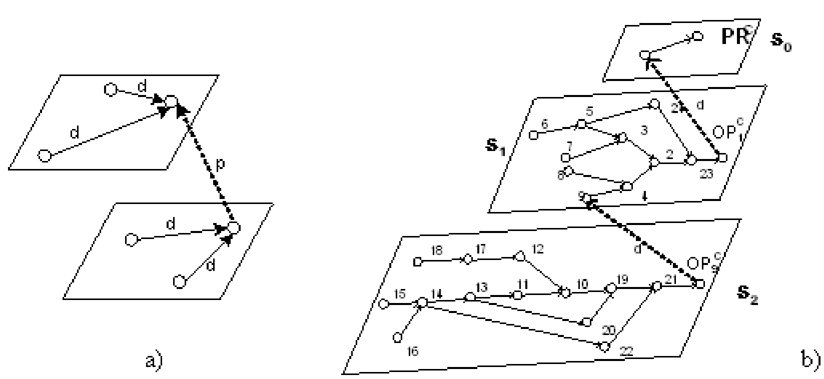


Figure 2. The second order S-tree by **p**-*relation (a) and the third order S-tree by* **d**-*relation (b)*



- A process, which has *UL=0*, is called *physical*: its *D* and *P* are real.
- *UL=1* corresponds to a *logical* process: its *D* and *P* have descriptions sufficient for their physical implementation.
- A *virtual* process has *UL=2*: its *D* and *P* exist only as mental images.
- *UL=3* is assigned to a *conditional* process (PR^C): its result has been declared but its *D* and *P* are presented by their symbols only.

Constructive proof of logical runability for PR^C consists in stepwise reduction of its *UL*. A step of reduction is referred to as *determination* of conditional, virtual or logical process. Each step of $PR^C=(D, P)$ determination is modeled by the search on appropriate sets for its *D* and *P*. Thus, each search assumes the run of two processes – *search for D* (*SD*) and *search for P* (*SP*):

$$SD \xrightarrow{d} \mathrm{PR}^{C} \xleftarrow{d} SP \qquad (2)$$

While the two-stroke determination of PR^C (Table 1), the objective of the *virtu al* (downward) determination is the reduction *UL=3→UL=2*; during the second or the stroke of *logical* (upward)

Table 1. Constructive proof of logical runability for a conditional process

Down stroke	**Up stroke**
$SD^V \xrightarrow{d} PR_1^C \xleftarrow{d} \mathrm{SP}^V$ **1**	PR_1^L =(D, P) **8**
PR_1^V $p\uparrow$ PR_2^c **2**	PR_1^L $p\uparrow$ PR_2^L $\Uparrow$ I^P **7**
PR_1^V $p\uparrow$ $SD^V \xrightarrow{d} PR_2^C \xleftarrow{d} \mathrm{SP}^V$ **3**	$SP^L \xrightarrow{d} PR \xleftarrow{d} \mathrm{SD}^L$ $p\uparrow$ PR_2^L $\Uparrow$ I^P **6**
PR_1^V $p\uparrow$ PR_2^V $\Uparrow$ I^P **4**	PR_1^V $p\uparrow$ $SP^L \xrightarrow{d} PR_2^V \xleftarrow{d} \mathrm{SD}^L$ $\Uparrow$ I^P **5**

determination, the reduction $UL=2 \rightarrow UL=1$ will take place.

This two-stroke cycle of PR^C determination results in *S*-tree (Figure 2*b*) where each node is the second order structure (2). Thus, while representing a complex process in *CPT*, the underlying principle is not a *D* decomposition with shaping a hierarchy of sub-processes, but making up a continuous structure out f process schemes.

Realization of Problems with Insolubility of the Second Kind

Let the structure (2) be elementary, i.e. when PR^C has a defined input (I^P) and does not need a generation of providing process:

$$SD \xrightarrow{d} \underset{I^{P\uparrow}}{PR^V} \xleftarrow{d} SP \qquad (3)$$

Reduce (3) to the form (4) assuming **d**-relation between the contents in brackets:

$$PrB = \langle SD, SP \rangle \xrightarrow{d} \langle \underset{I^{P\uparrow}}{PR^V} \rangle \qquad (4)$$

Following (Polya, 1965) while generic problem tackling, we distinguish a "*problem solution*" (the determined *D* and *P* from PR^L – the results of <*SD, SP*> performance) and an "*answer to the problem*" (the output of PR^L). The activity consisting in problem solving (<*SD, SP*> execution) and answer computing (PR^L run) is referred to as *problem realization*.

But expression (4) does not meet the structure precondition – the two processes in <*SD, SP*> are linked by empty relation. To restore for (4) the structure status, divide it into two schemes of processes where the relation between them is **d**-relation between their cores (the first scheme has no tail):

$$PrB = \langle SD \rangle \xrightarrow{d} \langle SP \xrightarrow{d} \underset{I^{P\uparrow}}{PR^V} \rangle \qquad (5)$$

Next, making **d**-relations implicit, we get two expressions named *problem schemes*:

$$PrB = \langle SD \rangle \langle SP, \underset{I^{P\uparrow}}{PR^V} \rangle$$

or (6)

$$PrB = \langle SP \rangle \langle SD, \underset{I^{P\uparrow}}{PR^V} \rangle$$

Thus, problem schemes like (6) get $n=2$ as each component scheme is the process structure with $n=1$. Incidentally, a structure made out of problem schemes (upon **p**-relation) should have $n=3$. Problem realization analysis in terms of problems schemes and their interrelations is named *problems schematics*. The use of this or that scheme from (6) depends on problem's specificity.

The problems that have both solution and answer are *soluble*. But, there are many more problems (posed usually directly upon a required result), which have an answer but have no solution. Those are referred to as insoluble of the *second kind*. (The problems that have neither solution nor answer have insolubility of the *first kind* – "perpetuum mobile" creation, for instance.) For all that, the answer availability to the problem with insolubility of the second kind indicates that instead of original (insoluble) some another problem is actually realized, the answer to which coincides with the answer to the initial problem. This another problem is said to be *conjugate problem* (*CP*) with respect to original one. Knowledge engineering gives the classic example of such a pair of problems: insoluble original – diagnosis problem, and soluble conjugate – knowledge inference problem. Answer deriving to a problem through realization of its *CP* is called *quasi*- or *q-realization* of the initial problem.

Turn now to design problem (*DP*), which has the reputation of ill-defined (Simon, 1970) or even wicked (Rittel & Webber, 1979), and write down its scheme:

$$P\ rB =< SD >< SP, PR^{V} >$$
$$\llcorner I^{P} =_{\text{needs and requirements}} \quad (7)$$

Any problem can be presented by "what is given" and "what is required". In case of *DP*, the "given" is a description of needs and a set of requirements, while the "required" is a sought-for product design. As the needs and requirements cannot be directly transformed into a goal design (the former and the latter belong to different conceptual worlds (Meijers, 2000), *DP* has no solution. Formally, there is no *D*∈*DPR*, so *DP* is a phantom problem. At the same time, an answer to *DP* (a design) is possible. Answer deriving to *DP* through realization of a conjugate problem (CP^{DP}) is referred to as *q*-realization of *DP*. Thus, the replacement, in respect to *DP*, of ambiguous "ill-defined" for constructive "insoluble of the second kind" means searching for CP^{DP}.

DESIGN THEORY PLATFORM: THE PARADIGM SIGNATURE IDENTIFICATION

A Conjugate Problem for a Design One

AXIOM 1

"There is no a priori DPR: DPR shaping comes to an end concurrently with design obtaining".

Let us begin the search for CP^{DP} with the search for a process conjugate to *DPR*∈*DP* and preliminarily named "not *DPR*" ("¬*DPR*"). Restore upon ¬*DPR* the scheme of CP^{DP}:

$$CP^{DP} =< SD >< SP, \neg DPR > \quad (8)$$

Virtual (or top-down, see Table 1) determination of ¬*DPR* by *D* requires generation of this or that virtual inputs for the process. For instance, any available design theory, which unavoidably employs some CP^{DP} by default, uses for the ¬*DPR* input a future product representation. The latter may have different levels of abstraction – a product during redesign or its abstract idea – and was called "context description" (Braha&Reich, 2003); such input refines ¬*DPR* to quasi-*DPR* (*qDPR*) or a process of product representation transformation. In this case the CP^{DP} problem scheme changes its view for (9) being ready for logical (bottom-up) determination of *qDPR*:

$$CP^{DP} =< SD >< SP, qDPR >$$
$$\llcorner I^{P} =_{\text{"contect description"}} \quad (9)$$

Explicit *qDPR* determination upon its object *D* (the result of *SD*>) is generally either impossible or restricted (*D* proposed in (VDI, 1987) or intended for Suh's model (1990, 2001)). Meanwhile, most of available design theories present this *D*∈*qDPR* as a "*DPR* model" (*qDPR*=(*D*, *P*)=(*"DPR model"*, *P*)), though inexistent entity, particularly *DPR*, cannot have a model by definition. Next, the performance of <*SP*> results in a "human master processor" (*H*) because "*DPR* model" realization is only within human's depth. Thus, completeness and quality of CP^{DP} solution depend on what is entering the ¬*DPR* input. How tear oneself away from the circle of mythic *DPR* modeling since a product representation could hardly be recognized as adequate input of ¬*DPR*?

Axiom 2

"Just as a physical product is the outcome of product design processing, so a product design has been an outcome of design process design implementation (processing)".

The stated axiom defines both generative process conjugate to *DPR* – call it *meta-DPR* or *mDPR* – and this process input, i.e. *DPR* design. *mDPR* should both generate *DPR* and produce a required product design through *DPR* performance. The new conjugate problem restored on *mDPR* is the problem of *DPR* design realization:

$$CP^{DP} =< SP >< SD, mDPR > \quad {}^{\Lsh} I^{P} = \text{DPR design} \tag{10}$$

It remains to construct *DPR* design and proceed with the search of solution to CP^{DP} when the latter is represented by the scheme (10). In the course of this problem solving, we have to identify a signature for the platform (systemic module) of Design theory. The solution per se is required to be triply independent: of a master processor, design task and design domain. We call such independence "3D independence (3Di). 3Di solution to CP^{DP} will be called also *systemic* solution. To obtain the latter, the main input of *mDPR* – *DPR* design – is required to be represented in 3Di form as well. *mDPR* platform or $mDPR_3$ will undertake the realization of 3Di *DPR* design. The input data of initial design problem serves for the second input of *mDPR*. We have not recalled it after (7) but it will serve for control data for the sought-for $D\in$ *mDPR*. Thereupon we refine the scheme of CP^{DP} under solution as in (11).

$$CP^{DP} =< SD >< SP, (D_{\in} mDPR, P_{\in} mDPR) > \quad \text{needs and requirements} = I^{P\Rsh} \quad {}^{\Lsh} I^{P} = \text{DPR design} \tag{11}$$

DPR Design Construction

Design Progress Concept

A product design does not emerge in a single-step action – it evolves over time. It is not desirable to have this evolving chaotic. Hence, any design process should realize some *design progress concept* (*DPC*). For instance, the famous model of engineering design process (Pahl & Beitz, 1984) mirrors *DPC*, which may be called stepwise refinement of design abstract description (conceptual, embodiment and detail design). Another example of *DPC* is the *evolution of population* (selection, mutation, recombination), presented in numerous papers on evolutionary design, e.g. (Vajna et al., 2005). For our project, we have taken the most natural, in our view, concept named *evolution of individual* – adaptation of the current product design state to a new state of *operation environment design* related to the sought-for product. This *DPC* is fixed in (10) by I^P-input of *mDPR*. In compliance with this *DPC*, the (operation) environment, which will accept a product after its physical implementation, is under design concurrently with a product. The initial state of environment is mirrored in (11) by I^D-input of *mDPR*.

Dynamic Design Representation

Thus, we should have to deal with representations of three designs – a product design, environment design and design process design. The idea of some unification of these representations (via a platform, for instance) knocks at the door, as they say. But yet more earlier by the same door, another idea stands in the wings – dynamic design representation, which is integral to design synthesis. Lastly, the all design representations within systemic module of Design theory, inclusive of the unified, should be 3D independent.

In point of a (physical) design itself, the listed concepts are unrealizable. Though the following axiom brings about a possible solution to the problem.

AXIOM 3

Just as a product is obtained via implementation of a product design, so a product design should be *obtained via implementation of product's design of design:*

design of design → design → artefact. (12)

An extension of string (12) to the left (design of design of design, etc.) is replaced with the sequence of design of design maturity levels (ML_i, $i=1,2,...$). Thus, the possibility of a unified, dynamic and 3Di representation is associated now not with a design but with a design of design or, as we shall call it in what follows, *logical design* (*L*-design) of a product, environment or design process.

We shall say that the sequence of *MLs* names reflects a *diachronic* (*dh*) or "historical" structure of *L*-design. Componentization of a single ML_i will be referred to as *synchronous* or *sh-structure* of a respective item from the *dh*-structure. At the same time, diachronic structure reflects only design *quasi-dynamics*, if synchronous representations of *dh*-items (description of componentization) are static, e.g. drawings. If we manage to impart dynamics to *sh*-representations, we shall obtain the corresponding representation for the *L*-design in whole.

Dynamics signifies a process. Hence, the case in point is a representation of each *dh*-structure item in terms of processes. This is possible if the content of any such item, from the very first one, should be the determination of the process scheme – namely, the scheme of *operation process* of the sought-for product. In that case, each *sh*-component will be specified by the structure of processes. This opportunity is provided by continuous process theory. Thus, with the assumption that designer perceives the future product in terms of its operation processes, we shall try to display his reasoning by:

1) shaping the sequence of *L*-design states (design *dh*-structuring),
2) dynamic representation of these states (by the structure of processes),
3) choosing the number of states that provides designer with a comfort design increment ("the quantum of evolution").

Successful virtual designing of a product comes to an end with deriving of a final *L*-design state – dynamic representation of the last item's (within $ML_1, \ldots, ML_n$) *sh*-structure. This sequence is given the name of *approximate model* (*AM*) of *L*-design. Incident ML_i and ML_{i+1}, $i=\overline{1, n-1}$, are linked with *embedding relation* ($\prec$).

Let us divide the *AM* into its *structure St* (*dh*-structure of *L*-design) and *semantics Sm* (the *sh*-content of each *dh*-structure item). If *Sm* of each ML_i is determined, we have *approximation model* of *L*-design: *AM=(St, Sm)*. Otherwise, it will be a *quasi-approximation* model – *qAM=(St, Sm*)*, where *Sm** is an abstract semantics. Then logical explicit designing is defined as *AM* construction for *L*-design or transformation of *qAM* into *AM* via refinement of *Sm**∈*qAM*. This *qAM* is called *L-design platform*.

Thus, the way to *DPR* design assumes construction of a unified *L*-design platform, refinement it till *L*-design and transformation the latter into a sought-for design. The role of objects to be unified will be performed by *dh*-structures of *L*-designs for a product and its operation environment, which design process should have to deal with. Those have to be obtained.

Primary dh-Structure for a Product L-Design

In compliance with the declared *DPC* (evolution of individual), we link by $\prec$-relation the following four product design states (*design goals*): *Prototype* (PRT), *Market version* (ITM), *Manufacturing version* (COM) and *Artefact* (ART).

$$PRT \prec ITM \prec COM \prec ART \quad (13)$$

To smooth the discrecity of design representation in terms of its states, split the development of each state into four stages. Each stage results in *design subgoal*: *quasi-system* or qSYS (a minimal set of product units that validates the concept of goal attainment), *system* or SYS (the extension of *q*SYS with control functions), *quasi-design* or *q*DES (space layout of the SYS constituents) and *design* or DES (the *q*DES components assigned with shape, materials, grades of finish and all necessary joints):

$$\text{qSYS} \prec \text{SYS} \prec \text{qDES} \prec \text{DES} \qquad (14)$$

Combine all design states into its diachronic representation – the data structure named *quasi-hierarchy* (*q*-hierarchy) – by closing the nesting hierarchy, i.e. making the latter actual across horizontal as well. The resulted *dh*-structure is shown in Figure 3.

Primary dh-Structure for Operation Environment L-Design

OE is a family of processes **{PR$_q$}**, which a product deals with throughout its lifecycle. There are only three kinds of relations between a product and a process from the family: a product can play the role of the *input*, *disturbance* or *processor* with respect to any process from **{PR$_q$}** (Figure 4*a*). Hence, we can break **{PR$_q$}** up to three constituent sets: {PR$_{q1}$}, {PR$_{q2}$} and {PR$_{q3}$} (Figure 4*b*).

Processes from {PR$_{q1}$} place on their input a number of *requirements* (*Rq*); this set is shared by the product life cycle stages (six, as in our case, or more). Processes from {PR$_{q2}$} impose *constraints* (*Cs*) on their disturbance. Processes from {PR$_{q3}$} specify for their processor the *operation conditions* (*Cn*). In turn, each constituent set is divided into motivated (we drop the details) number of subsets – the hierarchy in Figure 4*b*.

Figure 3. Primary dh-structure for a product L-design

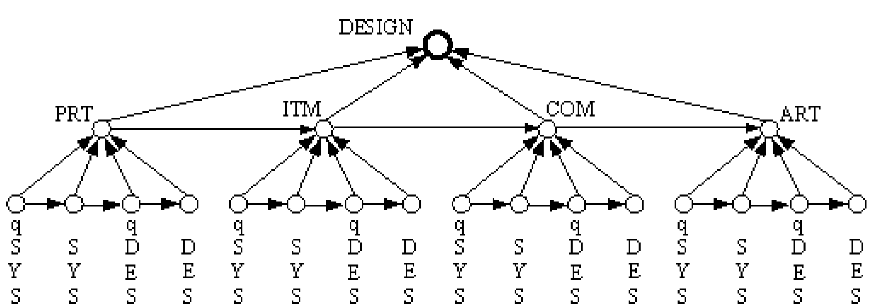
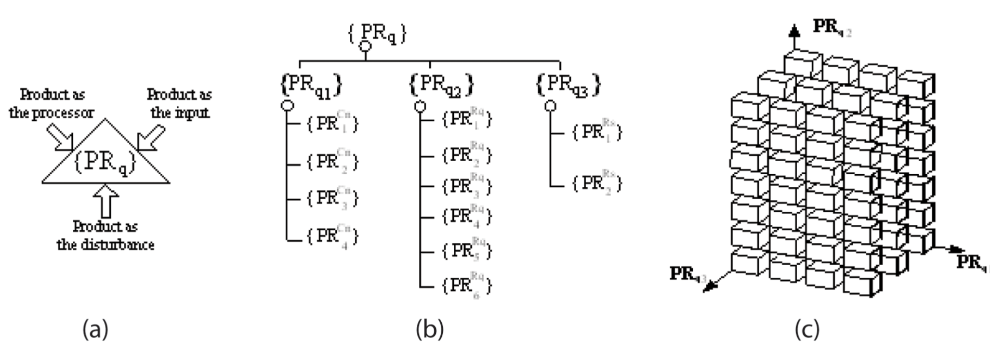


Figure 4. OE design dh-structure construction



As product-oriented *OE* should also be under design, it needs *P*-independent diachronic representation (hierarchies are *H*-oriented structures). Having taken the members of each set of processes (Figure 4*b*) for the vector of 3D space, we shape *dh*-structure of *OE* design. The latter is referred to as *&-cube* (Figure 4*c*).

Embedding relation links incident cells of &-cube along the *track* of its scanning. This track leads from the starting cell of &-cube – ($X_0Y_0Z_0$) – to its end cell – ($X_{max}Y_{max}Z_{max}$). Here are examples of tracks (brackets signify elements' iterations along the track up to their depletion):

- *(X(Z)(Y(Z)))* – layer-by-layer &-cube scanning along axis *Y*;
- *(X(Y)(Z))* – all *Z*-vectors, except the last one, are omitted in vertical layers (reduced layers);
- *(X)(Y)(Z)* – reduced &-cube (all *Y*-vectors are omitted, except the last one);
- *(X)(Z)(Y(Z))* – only the last vertical layer is taken wholly;
- *(X)(Z)(Y)(Z))* – the layer from the previous track is reduced.

Thus, the choice of &-cube scanning track defines both relevant power of {$\mathbf{PR_q}$} and variants of its structuring. These variants are also valid for the sets {Rq} and {Cs}, generated by {$\mathbf{PR_q}$} members (*OE* processes), and consistent with the concurrently derived design properties and characteristics. Incidentally, the obtained structuring for properties and characteristics is more insightful than in (Eder & Hosnedl, 2008).

From Design Platform to DPR Design.

The unified *dh*-structure is obtained by substitution of &-cube (Figure 4*c*) for terminals in the above *q*-hierarchy (Figure 3). After that, design *dh*-structure is valid for both product and *OE* designs, taking the status of $St \in qAM$. Whereupon, it takes only to refine $Sm^* \in qAM$ to come from *qAM* to approximate model of a product design (AM^P) and approximate model of *OE* design (AM^{OE}). The notion of design platform $qAM=(St, Sm^*)$ is completely pertinent for *DPR* design as well; the only difference is that "*L*-design" is used in this case without prefix because a physical representation for *DPR* design (a structure out of processes) coincides with the logical one.

Approximate model for *DPR* design is labelled as AM^{DPR}, and we come to it via refinement of $Sm^* \in qAM$. This refinement results in iterations of two processes (Figure 5) – synthesizing a new product design state (SPR^P, determination of the product operation process) and synthesizing a new operation environment design state (SPR^{OE}, changing for an increment the current set of conditions, requirements and restrictions, related to the sought-for product, and further paretization of this set).

Eventually, the obtained 3Di *DPR* design is represented by *AM*, in which every element of its *dh*-structure includes one iteration of the two processes mentioned above. Now when 3Di *DPR* design – the main CP^{DP} input (Figure 6) – has been obtained, we can proceed to CP^{DP} solving or determination of $mDPR \in CP^{DP}$ with respect to its subject (*P*) and object (*D*).

Figure 5. AM^{DPR} semantics

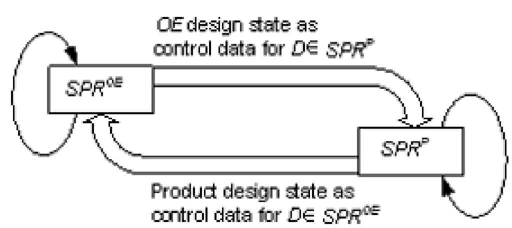


Figure 6. Input specification for the meta design process of 3Di

$$mDPR_3=(D, P)$$
P = 3Di *DPR* design

Systemic or 3Di solution to CP^{DP}

P-Solution to CP^{DP} or mDPR Determination on P

To get *P*-solution to CP^{DP} means to determine *P* from *mDPR=(D, P)*. Thus, the search for CP^{DP} solution on subject will proceed upon the expression (11) where <*SP*> means the search for *P*∈*mDPR*. To compile *P*, we have in general case: a) two types of processors – *H* and *C*; b) a list of members within a type; c) the kinds of relations between the types and members within a type (hierarchical subjection, cooperation and others).

We come to a virtual definition of the sought-for *P* in the following way. While *P*∈*PR* is a *physical* processor of some *PR*, *D*∈*PR* could be viewed as the *logical* processor with its own input – the processed control data (I^D). Then realization of any *D*∈*PR* is performed in general case by a pair of processors – a *master* one (P_w), which deals with the input of *P*∈*PR* (I^P), and *reference processor* (P_I), providing I^D for *D*∈*PR*. Let us consider the two as a single dual-core *P*. We call the latter a *diprocessor* (d*P*) – $P_w^{P^I}$. Diprocessor plays the role of action system (*AS*) platform, i.e. the base for shaping a real *P* for a particular *mDPR*. There are four variants for d*P* to be concretized: H^H, H^C, C^H or C^C. A process, which has the subject identical to one of the listed variants, will be referred to as *manual*, *computerized*, *automated* or *automatic* respectively.

Thus, CP^{DP} (virtual) solution on subject has been a d*P* with relation of interchangeability between *H* and *C*. This relation signifies that *P* of any type (*H* or *C*) is able to realize both work and reference function. Aiming at automated *mDPR*, we choose (logical solution) for the role of its *P* just d*P*=C^H. Then design automation gets under way from *mDPR* subject, and we have proposed instead today's intuitive design automation the more precise treatment of the notion. There are two ways of expanding d*P* to a real subject. First, each type of processors (*H* or *C*) has a number of instances that may be used. Secondly, P^I∈d*P* serves for the gateway to descendants whose action systems can take for the base any of the four d*P* types. Lastly, we assign to each member of d*P* the equal ability to perform both functions (master and reference). Thereby determination of *mDPR* upon subject gets the status of 1Di *P*-solution of CP^{DP}.

D-Solution to CP^{DP}

3Di *D*∈*mDPR* has been a procedure, which traverses the cells of &-cubes relevant to qAM^P and triggers in each of those the pair of processes (Figure 5). The corresponding meta-design process is designated as $mDPR_3$.

The Paradigm Signature for the Design Theory Platform

We call it Sg_1. The list of signature's paradigmants with their values is as follows:

1. Design Progress Concept = *evolution of individual*.
2. Conjugate Problem for a design one = *design process design implementation*.
3. Design (states) Representation = *dynamic*.
4. Fundamental definition of designing = *logical feedback synthesis*.

DESIGN PRACTICE PARADIGM IDENTIFICATION

Design process could not be other than domain-specific and task-oriented. At the same time, we retain it independent of a master processor type in d*P*, that is the aim of *mDPR* run is obtaining of 1Di design process.

In the light of the stated, the above obtained systemic solution to CP^{DP} (causing 3Di *DPR*) is out of practical sense as 3Di *mDPR* cannot transform entering its input 3Di *DPR* design into 1Di *DPR*.

To get 1Di *DPR* as a result of CP^{DP} realization, we need the lDi status either for *mDPR* object and subject or for its I^P-input – *DPR* design. The choice of this or that variant will be defined by a strategy of design system construction. We shall dwell on the variant of successive concretization of 3Di solution to CP^{DP} that is deriving 1Di solution from a preliminary obtained 2Di solution. The corresponding *mDPR*s will be denoted as $mDPR_1$ and $mDPR_2$ respectively.

2Di Realization of the Conjugate Problem

2Di D-Solution to CP^DP^

This solution consists in determination of $mDPR_2$ upon its subject (*P*) and object (*D*). Given the problem scheme – CP^{DP}=<SP><SD, $mDPR_2$> – let us feed the $mDPR_2$ input (Figure 7) with *DPR* design approximate model (AM^{DPR}) and design task (a sought-for product description and design goal).

CP^{DP} solution on subject, i.e. $P \in mDPR_2$, has been the above obtained *AS*. CP^{DP} solution on object (*D*) is a procedure that will allow to adjust the qAM^P structure to design goal and design subgoals. This adjustment consists in: i) selection of a required design goal (PRT, ITM, COM or ART) and terminal subgoal (qSYS, SYS, qDES or DES) for each inclusive goal; ii) elimination of irrelevant life cycle phases in subgoal's &-cube; iii) assignment of &-cube scanning path. Just the appropriate meta design process, designated as $mDPR_2$ with 2Di object (*D*) and 1Di *AS* in a subject (*P*) capacity, is taken as relevant for systemic module of design theory.

Figure 7. Input specification for the meta design process of 2Di

$$mDPR_2 = (D, P)$$

design task $= I^D$ ↰ ↳ $I^P = AM^{DPR}$

2Di CP^DP^ Realization

We assume here a logical realization of CP^{DP} – making arrangements for $mDPR_2$ rendering. The facilities for $mDPR_2$ implementation (a special purpose OS) have gotten the name of domain-independent *design machine* (Sedenkov & Guziuk, 2004).

1Di Realization of the Conjugate Problem: Design system

To obtain 1Di solution to CP^{DP}, there is a need for information about design domain. In the general case, such information is represented by a collection of domain-specific procedural and declarative knowledge: domain-specific SW libraries, theories, gained experience, design methods, etc. Acquisition, representation and assignment of such knowledge are the functions of the third module of Design theory, which gets the name of Domain Knowledge Management System (*DKMS*). A complete study of this module is beyond the scope of this chapter, while the outline of its signature looks like: 1) knowledge structuring; 2) knowledge storing; 3) knowledge access.

Starting 1Di solving of CP^{DP} (determination subject and object of $mDPR_1$), we claim the following input to $mDPR_1$ (Figure 8).

Then we take *P=DM* to be solution to CP^{DP} on subject, while the problem solution on object is taken of a procedure that couples in each &-cube's cell from qAM^P and qAM^{OE} structures the pair of processes (Figure 8) with *DKMS*.

Figure 8. Input specification for the meta design process of 1Di

$$mDPR_1 = (D, AS)$$

$DKMS = I^D$ ↰ ↳ $I^P = qAM^P \,\&\, qAM^{OE}$

The Paradigm Signature for Design Practice

Paradigm signature of the second design theory module (Sg_2) consists of the following three paradigmants:

1. Applied definition of designing = *transformation of design platform (qAM) into design.*
2. *mDPR$_1$* = (Subject: *DM*, Object: *The procedure coupling DM with domain knowledge*).
3. Design system architecture = *Platform (DM) based Design System*.

SUMMARY: THE OUTCOMES OF NEW DESIGN PARADIGM REVEALING

Design Theory Rearrangement

In search of a new Design practice paradigm, it was imposs ible to avoid to refine some basic aspects of Design theory – design problem and process, design representation, constructive definition of Design. Available Design theories do not consider these issues in their close relationship. At the same time each theory gives its own definition of Design, that is claims a conjugate problem to solve. The adequacy of this problem is called by Design practice in question, as a rule. Besides, the proposed theories virtually do not take into account the participation in designing the formal processor (computer) though its expansion into the process will only grow. The next challenge is the steady increase of design process complexity. In sum, these circumstances cause considerable confusion, discordant opinions and chaotic tendencies in research (Hundal, 1990; Tomiyama, 1990). To take this situation for the reference point in search of adequate Design paradigm was as recklessly as to propose a regular new Design theory.

We decided to rearrange the current design theory through partition it into three conditionally independent units and their content revision. The unit of Design theory basic principles was named *systemic module* or a platform of the theory. Solutions to the problems related to this module are required to be independent of domain, design task and a type of processor – *C* or *H*. Systemic module is followed by a module of applied design aspects where the solutions produced at systemic level are coupled with design domain and design task. In turn, the applied module initiates the third module that provides designing with procedural and declarative knowledge. Each module was attached with the established signature of its paradigm.

Incidentally, the auxiliary (modular) concept of design theory reorganization has proved to be tempting as the concept of its initial arrangement as well. It seems a promissory alternative to both attempts to develop a General Design theory and tendency to have specific theory in every design field. Modular approach works in the best way also for theory consolidation, the calls for which sound ever more insistently (Birkhofer, 2006). This approach define three main lines of the theory development – fundamental, applied and design knowledgemation aspects – and distinct directions (upon signatures' paradigmants) within the bounds of each module. The relative autonomy of modules enables them to progress independently and at own rate.

In the course of systemic module establishment (design theory platform), we have considered the main, in our view, delusions ingrained in engineering design – design problem, design process, and product representation. The first among those is the design problem. It does not exist as a soluble problem. Instead of *DP* there is realized a conjugate problem (quasi-realization of *DP*). The choice of the latter will define the paradigm of the applied module, as well.

Design problem quasi-realization crosses the positivism of H. Simon (1970, designing as problem solving) and constructivism of D. Schön (1983; 1992, the study of designer's activity). In addition to the Simon's concept, we have changed the coined characteristic of design problem ("ill-defined") for the more promising one – "un-solvable of the second kind". This resulted in realization of the conjugate problem relative to the phantom *DP* – the problem of design process design implementation. In addition to the Schön's concept, we have proposed the proactive model of interaction between a product design and product operation environment ("evolution of individual") embodied in design process design.

The next modification was concerned with a design. Instead of misleading but widely discussed *product* representation (Ariyo et all., 2008), we have proposed continuous (dynamic) representation of a *product design*. This made possible explicit structure synthesis of both designs process and product.

It was now a design process turn. The latter a priori does not exist though it has been the current focus of design research community. Within the limits of Continuous Process Theory and its pivotal formalism – process scheme *PR=(D, P)* – the notion "(design) process model" becomes pointless. There are levels of uncertainty introduced for a process, pursuant to which it can be conditional, virtual, logical or physical. Hence, *DPR* model (Pahl & Beitz, 1988) has been an abstract procedure in the scheme of virtual process – *DPR=(DPR model, P)*, the refinement of which comes to end only when designing is over. In reality, there is in operation just a meta design process, which composes the design process proper part-by-part, supports these parts performance and, eventually, provides through this performance a required design. 3D independent meta-*DPR* (*mDPR* platform) is also defined within systemic module of design theory. Then a unified design platform had been constructed that enabled to represent design process design entering meta-*DPR*.

Thus, renunciation of the listed delusions shifts the focus from design problem to its conjugate problem, from design process to meta-design process, and from representation of a product to representation of a design – particularly continuous representation. But the all denoted innovations within systemic module would have been out of place without the proposed formal tools – Continuous Process Theory. While other design theories address to external facilities (which entail external problem as well) – for instance, issues in set theory or logic (Tomiyama & Yoshikawa, 1987; Hatchuel & Weil, 2003), we have used the ad hoc tooling. *CPT* meets a lack of scientific principles in engineering design and shifts the latter to a discipline in its own right.

The Impact of the New Design Paradigm on Design Practice

Design paradigm signature consists of three paradigmants. The names and values of the new one in comparison with the current design paradigm (based on German systematic model (Pahl & Beitz, 1984) are given in Table 2.

Within the new Design paradigm, *mDPR* deals with not a model but design of *DPR*, realizing it initially at systemic and then at applied level in line with constructive definition of Design for each theory's level – "Design as *DPR* design implementation" and "Design as transformation of design platform into design" correspondingly. Thus, with the account of Axiom 1, the quotation "Most researches and practitioners recognize that design process cannot be fully formalized mathematically" (Braha & Reich, 2003) gets a new sense: there is no such a necessity. Henceforth, exactly $mDPR_i$ ($i=1,2,3$) is intended to attract main activities directed to resulted *DPR* perfecting.

In whole, new Design paradigm paves the way to new technologies in design computerization, design automation and design knowledgemation. For instance, the use of *design system platform* (design machine, *DM*) leads to the new efficient and effective technology of design computerization (automation): the compilation a wind range of domain-, product-, user-, and media-oriented design systems, via replication of domain-indepen-

Table 2. The comparison of two design paradigm signatures

Paradigmants **Design Practice Paradigm**	**Conjugate Problem under solution**	**Conjugate Problem Solution** (*mDPR* input, its subject and object)	**The system that supports *mDPR* run** (Design System)
Current	Stepwise refinement of a prototype abstract description.	I^P = *(prototype, conceptual, embodiment) design description.* I^D = *design task.* *mDPR* subject = H^C. *mDPR* object = = *substantially implicit.*	Whole-cut and high-value item with long-term development but short durability. Supports Computer Aided Design (CAD).
New	Transformation of design platform into a product design.	$I^P = qAM^P \& qAM^{OE}$. I^D= *DKMS.* *mDPR* subject= *DM.* *mDPR* object= *D that supports interface between DM and DKMS.*	Platform based, holistic and product (domain, user, media)-oriented. Supports Computer Urged Design (CUD).

dent *DM*, and further extension of each specimen with domain-specific knowledge (design knowledgemation) via design knowledge management system. Design knowledge structuring within *DKMS* is guided by the structure of the unified design platform (*qAM*).

Due to *P*-independent theory of the systemic module, it is clearly stated the partner but not merely instrumental role of computer in Design: action system of $mDPR_1$ employs C^H platform, whereby we may identify the new paradigm as *Computer Urged Design* (CUD) that replaces CAD.

And one more benefit of the new design paradigm: it eliminates the problem of design process complexity: the realized $mDPR_1$ does not critically depends on complexity of a product under design.

CONCLUDING REMARKS

Any design theory should begin with propositions concerning the way of unavoidable design problem (*DP*) realization. To this end, it should have not involved but an ad hoc scientific base. In our case, the role of the latter is performed by Continuous Process Theory and Problem Theory.

(As for available design theories, most of those avoid to focus on design problem as such and begin just with design process modeling. The appropriate tooling is borrowed from other disciplines. This results in a doubtful resolving power of the tools and some extra problems for designing borrowed from the donor disciplines together with tools.)

Insolubility of design problem validated by *CPT* entails the search for a conjugate problem (CP^{DP}). For the role of CP^{DP}, Axiom 2 has identified the problem of design process design implementation. The process of such implementation was called *meta-DPR* (*mDPR*). *mDPR* dynamically generates design process, supports it performance and concurrently becomes the part of the process.

(Available design theories also realizes a conjugate problem but does this by default. Mostly it is one and the same problem of transformation a future product representation, which may have different level of abstraction. Thus, design theories with such CP^{DP} differ only in their input representation, involved tooling for the input transformation, and appropriate procedure. Schematic or concrete description of the latter is inconsistently called "*DPR* model", though inexistent entity, in particular *DPR*, cannot have a model – the only primary description of inexistent entity is a design. Due to the common CP^{DP}, these theories share such characteristics as catering for adaptation design, indispensable availability of a prototype, the luck of a structure synthesis problem solution, an obscure way of design automation, inability to manage the growth of design process complexity, and exclusive *H*-orientation as well. The latter prevents efficient usage in design a computer, reserving for it the role of instrument on demand. And though theory's human orientation is natural, we propose the trade-off decision – to have in the three-module theory structure a processor-independent systemic module.

Next, available design theories used to deal rather with *product models* than design representations. As inexistent entity cannot have a model, we considered the way of *product design* representation, particularly *dynamic* design representation required for explicit design synthesis. When designing, we have to deal with three designs – a product design, its operation environment design, and *DPR* design. So, the partial unification of their representations seemed both necessary and beneficial. The result of such unification has manifested itself in *design platform*. The latter gave rise to constructive definition of designing in the new paradigm: successive semantic attribution of each design platform cell.

Generally, the platform concept is also in the focus of the new design paradigm. Except design platform, it has defined generative *DPR* platform (*mDPR*), action system platform (diprocessor, d*P*), design system platform (design machine, *DM*) and design theory platform (systemic module). Besides, the bulk of facilities used for the platform technology description may shape one more platform – *design language platform*, which should allow to change from the current substantially intuitive notion base of designing to the more rigor one and thereby to serve for design coordination over distance and across professions. In whole, the platform-based reasoning paves the way for domain-independent approach to theory and research in design.

And one more point is noteworthy. There are an impressive number of questions, induced over time within design discipline, that still remain without answers (Sedenkov, 2008). The modular structure of Design Theory with identified signatures of modules' paradigms, the base of scientific principles in the form of *CPT* and platform-based technology enable to field these questions.

The main benefit of the new design paradigm for education is that it makes the primary design related processes ($mDPR_i$) observable, teachable and learnable. Incidentally, a pilot version of design system platform (design machine) is already in use for students training.

In whole, development of design practice paradigm contributes to shaping Design as "a coherent discipline of study in its own right, based on the view that design has its own things to know and its own ways of knowing them" (Cross, 2007, p. 3).

REFERENCES

Ariyo, O. O., Eckert, C. M., & Clarkson, P. J. (2008). Hierarchical decompositions for complex product representation. In *Proceedings of International Design Conference – DESIGN 2008* (pp. 737-744). Dubrovnik.

Birkhofer, H. (2006). The consolidation of design science – a critical review of the status and some proposals to improve it In Vanek, V., & Hosnedl, S. (Eds.), *AEDS 2006 Workshop* (pp. 13–22). Pilzen, Czech Republic: University of West Bohemia.

Braha, D., & Reich, Y. (2003). Topological structures for modeling engineering design processes. *Research in Engineering Design*, *14*, 185–199. doi:10.1007/s00163-003-0035-3

Cross, N. (2002, February). *Design as a discipline*. Paper presented to assist discussion at the conference "The Inter-disciplinary Design Quandary", De Montfort University, Leicester, UK.

Cross, N. (2007). Forty years of design research. *Design Studies*, *28*(1), 1–4. doi:10.1016/j.destud.2006.11.004

Dixon, J. R. (1988). On research methodology towards a scientific theory of engineering design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *1*(3), 145–157. doi:10.1017/S0890060400000251

Eder, W., & Hosnedl, S. (2008). *Design Engineering, A Manual for Enhanced Creativity*. Florida, USA: CRC Press.

Hatchuel, A., & Weil, B. (2003). A new approach of innovative design: an introduction to C-K theory. In *Proceedings of the International Conference on Engineering Design – ICED'03* (pp. 109-110, executive summary). On CD-ROM. Stockholm.

Hundal, M. (1990). Research in design theory and methodology in west Germany. In *Proceedings of the ASME Design Theory and Methodology Conference – DTM'90* (pp. 235-238), Chicago. New York: ASME.

Lossack, R., & Grabowski, H. (2000). The axiomatic approach in the universal design theory. First International Conference on Axiomatic Design. Cambridge, MA.

Meijers, A. W. M. (2000). The relational ontology of technical artifacts In Kroes, P. K., & Meijers, A. W. M. (Eds.), *The Empirical Turn in the philosophy of technology*. Oxford: Elsevier Science.

Pahl, G., & Beitz, W. (1984). *Engineering Design*. London: Springer-Verlag.

Polya, G. (1965). *Mathematical Discovery. (VII)*. New York, London: John Wiley&Sons.

Rittel, H., & Webber, M. (1979). Dilemmas in a General Theory of Planning. *Policy Sciences*, *4*, 155–169. doi:10.1007/BF01405730

Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.

Schön, D. A. (1992). Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems*, *5*(1), 3–14. doi:10.1016/0950-7051(92)90020-G

Sedenkov, V. (2008). An attempt to answer perennial design questions In Rohatynski, R. (Ed.), *Design Methods for Industrial Practice* (pp. 21–30). Poland, Gdynia-Zielona Gora.

Sedenkov, V. (2008). Design problem *q*-realization and its outcomes In Vanek, V., Hosnedl, S., & Bartak, J. (Eds.), *AEDS 2008 Workshop* (pp. 91–98). Pilzen, Czech Republic: University of West Bohemia.

Sedenkov, V., & Guziuk, Z. (2004). Design machine: theory and implementation. In I. Horvath & P. Xirouchakis (Eds), *Tools and Methods of Competitive Engineering* (Vol. 2), (pp. 1119-1120, executive summary). (On CD-ROM). Rotterdam: Millpress.

Simon, N. A. (1970). *The Science of the Artificial*. Cambridge, UK: MIT Press.

Suh, N. P. (1990). *The Principles of Design*. New York: Oxford University Press.

Suh, N. P. (2001). *Axiomatic Design*. New York: Oxford University Press.

Tomiyama, T. (1990). Engineering design research in Japan. In *Proceedings of the ASME Design Theory and Methodology Conference – DTM'90* (pp. 219-224), Chicago: ASME.

Tomiyama, T., & Yoshikawa, H. (1987). *Extended General Design Theory*. In Design theory for CAD. North-Holland: Elsevier Science Publishers B. V.

Vajna, S., Clement, S., Jordan, A., & Bercsey, T. (2005). The Autogenetic Design Theory: An Evolutionary view of the Design Process. *Journal of Engineering Design*, *16*(4), 423–440. doi:10.1080/09544820500267781

VDI-Guideline 2221/1987: Systematic Approach to the Design of Technical Systems and Products. VDI, Düsseldorf.

Yoshikawa, H. (1981). General Design Theory. In T. Sata & E. Warman (Eds), The IFIP WG 5.2-5.3 Working Conference (pp. 35-57). North Holland, Amsterdam.

KEY TERMS AND DEFINITIONS

Meta Design Process (*mDPR*): The process intended to implement design process design. Through this implementation, *mDPR* dynamically generates design process, supports its performance and supplies the final product design state. Splitting a virtual design process into the real *mDPR* and a pair of synthesis processes (for the product and environment designs) has proved to be a promising deed.

Paradigm Signature (*Sg*): A number of paradigmants sufficient for a unique identification of a paradigm.

Paradigmant: Distinguishing feature of the paradigm descriptive of its certain property.

Problem Quasi-Realization: A situation when instead of initial (insoluble) problem another problem (conjugate) is under realization. The answer to the conjugate problem coincides with the answer to initial one.

Problem Scheme: The structure of the second order made up of process schemes names by relation of determination: *PrB= <SD><SP, PR>*. Each delimiter (angle brackets or comma) indicates **d**-relation (from left to right) by default. *SD* – search for *D*, *SP* – search for *P*, *PR* – performance *D* by *P*.

Process scheme: A simple formalization for a process *PR* intended to represent the latter by the pair – a procedure *D* and processor *P*: *PR= (D, P)*. *D* describes the function of *P* over its input I^P. Process scheme is a basic unit for shaping process scheme structures – continuous representations of complex processes. To this end, a set of process schemes is provided with two initiating structure relations (**p**- and **d**-relation).

Quasi Design Process (*qDPR*): The process intended to transform some initial representation of a future product design into a final state of the design.

This work was previously published in Handbook of Research on Trends in Product Design and Development: Technological and Organizational Perspectives, edited by Arlindo Silva and Ricardo Simoes, pp. 18-38, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 85
Dynamics in Knowledge

Shigeki Sugiyama
University of Gifu, Japan

ABSTRACT

Since the idea of "artificial intelligence with knowledge" had been introduced, so many thoughts, theories, and ideas in various fields of engineering, science, geology, social study, economics, and management methods have been proposed. Those things have been started as an extension of modern engineering control theories and practices. Firstly, expert system by using IF-Then rules came up to at a production spot in manufacturing, and then agent system method by using intelligent software programs for design, planning, scheduling, production, and management in manufacturing. And then after, the idea of "Knowledge" burst into the artificial intelligence field as a real aid for getting any purpose to be accomplished by having augmented the past key knowledge in terms of management (controlling). However, those augmented knowledge methods used to have usages only in a limited small area. In addition to this, lots of works have to be done before making the systems work for a target problem solving. And what is worse, lots of parts of systems have to be customized for a new application. This chapter introduces a new direction and a method in "Knowledge" by inaugurating the brand new idea of "Dynamics in Knowledge," which will behave more flexibly and intelligently in real usages.

INTRODUCTION

As an introduction, here briefly touches upon the conventional methods for engineering controls, an expert system, an agent method in term of management, which will be related with the future advancements.

DOI: 10.4018/978-1-4666-1945-6.ch085

History of Management Method in Control

One of the earliest open-loop control systems was seen by Hero's device for opening and closing the doors of a temple in the first century at Greece. The opening and closing required inputs to the system were by lighting and extinguishing of the fire of the altar. By the expanding the hot air under the fire made the water transfer from the container into the bucket. As the result of this, the bucket became heavier because of the incremented and fulfilled water, so that the bucket naturally caused to descend because of its gravity. And so its movement made the door spindles turn by means of a rope, and which could open the door. On the other hand, the door could be closed by moving the water of the bucket back to the container again by cooling the container with extinguishing the fire of the altar so as the pressure was being reduced. Thus the bucket became lighter, and then the bucket was moved down so as the door was closing.

The first significant work in automatic control was James Watt's centrifugal governor for the speed control of a steam engine by controlling the vapor pressure of the tank in stable.

And the frequency-response method made it possible for engineers to design linear feedback control systems that could satisfy lots of the engineering requirements. And then, the root-locus method in control system design was developed, which was working more intimately.

These methods are the major control methods. And then, because of necessity to treat many inputs and outputs, these methods became less significant. And then modern control theories have been developed. In modern control theories, linear control method, non-linear control method, and discrete control method have been developed by using a computer more and more even for complex systems.

After the classical control methods had been developed, the application of modern control theories expanded quite rapidly even in the fields of geology, economics, medicine, and sociology, etc. What is more, with developing of the micro processors process speeds and those software, it has become possible to have a discrete control and an expert system which is able to mimic human skills in manufacturing and management (thought of thinking).

Almost in the same period, Petri Net, Neural Networks, Fuzzy Logic by L.A. Zadeh, GA, Immune Control Method, Chaos, Complexity theory, and others have come up to the world as Intelligent Management Methods. And nowadays these methods are tried to be used in more complex and huge systems as a total management system.

PAST, PRESENT, AND FUTURE SITUATIONS IN PRODUCTION AND MANAGEMENT

Technologically Speaking, It has the Following History and the Future

1. One skilled person production
2. Power assisted by the nature (water) that helped a skilled person production
3. Power aided by artifacts that promoted a skilled person's production
4. Jobs Distributed System for Production
5. Worker Assisted by Mechanics Production for a tiny spot
6. Worker Assisted by Mechanics Production for a spot
7. Worker Assisted by Mechanics Production for an area
8. Worker Assisted by Machine Production
9. Worker Assisted by Mechatronics Production
10. Worker Assisted by Semi-Automatic Machine Production
11. Worker Assisted by Automatic Machine Production
12. Worker Assisted by Robot Production
13. Worker Assisted by Expert System
14. Manager/Worker Assisted by Artificial Intelligent System, (at present.)
15. Manager Assisted by Information Technology and Artificial Intelligent Management System, (in the future.)
16. Manager Assisted by Artificial Intelligent and Information Technology Management System

17. Manager Assisted by Dynamic Knowledge Base System
18. Semi-Autonomous Artificial Intelligence Management System
19. Autonomous Artificial Intelligence and Knowledge Management System
20. Management System assisted by Autonomous Artificial Intelligence

In these thirty years, especially, we had the drastic change in production and management. That is to say, the assisted parts in production have shifted from Power to Human Knowledge as shown below.

Power Assisted Skill → Replace Part of Skill by Knowledge → Replace Part of Job by Knowledge → Job Assisted by Knowledge → Human Assisted by Knowledge.

These phenomenon are observed by the facilities that we have been using now as shown below.

1. Exclusive Use of Computer Assisted Machine
2. Group Controlled Machine System
3. Semi-Automatic Machine and Automatic Machine
4. AGV
5. Robot
6. CIM
7. Process Control by GA, Fuzzy Logic, Neural Network, and Artificial Intelligence
8. Total Control and Management System
9. Knowledge Base
10. Management System in Complexity
11. Data Mining (Sort)
12. Total Management System for various fields

Through observing these developments in production (management) and in facilities, it is obvious to say that many kinds of production and management systems and facilities including computer have been developed in order to assist a manager/worker, physically, in an effective way. And now it is a small and tiny phenomenon, but Human Assisted System by using Knowledge Base or Artificial Intelligence with Knowledge is beginning to come up for real use.

Problems in General

Here, the general phenomenons in the management systems are discussed in terms of problems.

Because of the social needs/changes or because of the restrictions in the production and the management methods, we have the major four key difficulties and problems as mentioned below.

1. **Agile Production/Management (AP):** Firstly, in general, the most important thing for industries to be existed is to offer desired production goods at the places wanted, on the time desired with the highest quality and the necessary quantity. This is the only way to satisfy Need (Desire) in a society (Sugiyama (Ed) et al. 2002c). There are many kinds of facilities and many kinds of management systems that will facilitate these kinds of productions and managements quite easily, but these are the things that money can buy. So under this situation, those can be bought from anywhere in the world if there is a greedy person to do so. This means that those management systems can be constructed if there is a person to do so with money. In another words, once a good system is invented, there must be someone who is going to use the same system for lowering a cost production at somewhere in the world. So the companies will have been always facing this kind of the problem (Sugiyama (Ed) et al. 2002c) that is never ignored.
2. **Huge and Complex System (HC):** Secondly, in order to improve productivity and efficiency in a total production and management, the control areas of fields have been expanding more and more from a tiny spot of a machine to a whole company. So the system to be treated will become so huge

and complex. After all, the conventional methods in a production and management system will have become less effective in sense of getting the profits.

3. **Networking Allover (NA):** Thirdly, management in production and in sales, the globalization comes up, so that the functions of a company have been distributed at various places in the world geologically, but those distributed places always need very intimate and close relationships one another in order to create the most profitable output. So they need to behave as if they are very close and near geologically, in another words, they need to behave as if they are one unit of function.
4. **Intelligence like Human (IH):** Fourthly, it is quite important to watch the system closely whether it goes well or wrong even though a system is controlled automatically. Because a system easily goes wrong by a disturbance or by miss usage of systems. And to get it back to the normal situation it usually needs the proficiency of human to do. But ideally it would be the best if a system can do them intelligently. So in order to reduce this burden by human, the system needs a kind of intelligence to do this autonomously in replace of human.

RELATED WORK AND DIFFICULTY

Here examines how the today's methods including the conventional methods work and are working in order to reduce the problems (AP, HC, NA, IH) mentioned above. And also if today's methods do not work well, here shows what sort of difficulties they are facing.

Conventional Method

For controlling a motor, a machine, a transportation, etc., we have a method of continuous control method. But as every behavior of the control targets has to be written mathematically in detail, it is not easy and tough thing to do. As the result of this difficulty, the control areas are very limited within the special fields. If it is failed to express the target system mathematically in detail, lots of optimization have to be made. And as the result of this, it may cause to loose the accurate control.

In the early days of controlling and management, we have the following types of control methods, which are Open-loop control system and Closed-loop system. Open-loop control system (Bertalanffy 1950) is the one in which the control action is not influenced by the results of that action. Some desired or reference input is sensed and amplified and then operate on by the controller.

So this control method has a difficulty in that Open-loop control can be used in practice only if the relationship between the input and output is known and if internal and external disturbances are not much concerned.

A closed-loop control system (Saucedo et al. 1968) is the one in which the output quantity is in feedback for comparison with the input, and the difference, or the error quantity, is applied to the amplifier in order to reduce the error and it cause to bring the output of the system to a expected value. This is better off than the open-loop control system in terms of the accuracy, and what is more, this system is more robust under the disturbance.

Mathematically, generally speaking, the above two control methods can be shown in the simplest expression as shown below.

$[O] = [G] \times [I]$ (1),

where,

[O]: Output,
[I]: Input,
[G]: Transfer Function.

So as the results of these discussions, it is quite obvious to say that, for applying these methods, the target object of the whole system needs to be defined mathematically with continuous functions (linear or non-linear functions). That is to say, however large and complex the target system is, it should be clearly defined mathematically all through the system. And once the system has been defined, it should keep the system structure whatever happens. What is more, the variable numbers to Input are limited to certain values, that is to say, we cannot input arbitrarily number of variables.

As the results of these, we face the following difficulties;

1. It is hard to apply these methods in an excessively complex and huge system.
2. It is quite hard to use an irrational function.
3. It is impossible to use a graphically and experimentally expressed data.
4. It is hard to put an intelligence into the system.
5. It is quite hard to modify the system to the continuous change of the conditions.

So the problems of (AP), (HC), (NA), and (IH) are left as the difficulties.

Discrete Control Method

Discrete-time systems, or sampled-data systems (Saucedo et al. 1968), are dynamic systems in which one or more variables can change only at discrete instants of time. These instants may be the performing time at which some physical measurement is performed or the time at which the memory of a digital computer is read out, etc. The time interval between two discrete instants is taken to be sufficiently short so that the data for the time between these discrete instants can be approximated by simple interpolation. Discrete-time systems differ from continuous-time ones in that the signals for a discrete-time are in sampled-data form, so that the basic idea is almost the same as the methods mentioned above section, but only the mathematical expression is different.

Mathematically, generally speaking, the above control method can be shown in the simplest expression as shown in the equation below.

[O]* = [G]* x [I]* (2),

Where,

[O]*: Discrete Output,
[I]*: Discrete Input,
[G]*: Discrete Transfer Function.

Some of the advantages of digital method comparing with very conventional analogue ones may be summarized as follows.

1. Digital method is able to perform complex computing with a constant accuracy at high speed.
2. Digital method makes possible to use a larger system compared with the analogue method.
3. Digital method is very flexible in applications.

So as the results of these, this method is better than the conventional methods in a sense of treatment in the system operation, however there are almost all the problems of (AP), (HC), and (IH) that the other conventional methods have.

BACKGROUND

Artificial Intelligence Method

Artificial intelligence is a growing field that covers many disciplines as stated by Nick Cercone and Gordon MacCalla in the paper "Artificial Intelligence."

Artificial intelligence is an interdisciplinary field. It is related with Artificial Intelligence derives, including contributions from psychology, *philosophy, linguistics, electrical engineering, and computer science. The intersection between psychology and Artificial Intelligence centers on the sub areas known as cognitive psychology and psycholinguistics. Philosophy and Artificial Intelligence come together in the areas of logic, philosophy of language, and philosophy of mind. Intersections with linguistics include computational linguistics, psycho-linguistic. Philosophy and Artificial Intelligence come together in the areas of logic, philosophy of language, and philosophy of mind. Intersections with linguistics, include computational linguistics, psycholinguistics, and socio-linguistics. Mutual concentrations between electrical engineering and Artificial Intelligence include image processing, pattern recognition, and robotics. Computer science overlaps Artificial Intelligence proper and the related field of adaptive system.*

So there are so many ways to define the field of Artificial Intelligence as Nick Cercone and Gordon MacCalla are mentioning. That is to say, as a science, it will be essentially part of cognitive Science, and the goal of Artificial Intelligence is to understand principles that will make intelligence possible to work. As a technology and as a part of computer science, the goal of Artificial Intelligence will be to design intelligent computer systems that will show the characteristics associated with intelligence in human behaviour; that is to say, understanding language, environment, behaviour, learning, reasoning, solving problems, and so on. Especially Artificial Intelligence including knowledge representations, problem solving, learning, natural-language understanding, computer vision, robotics, Artificial Intelligence languages, expert systems, and several others is for studies. Many of these areas are related to one another theoretically and technically.

Knowledge Representation and a fundamental element of Artificial Intelligence were how to express information of a target object, a treatment in action, and processes into data structures and procedures. The most popular current approaches are rules, definitions, and basic logics without inferring, and production mechanisms. Knowledge representation schemes can be classified into declarative and procedural ones. In a case of a declarative one, it generally requires wasteful searching of data. And a procedural one base is relevant with programming methods.

From here on it is described a little bit more in detail about the typical areas of researches related with Artificial Intelligence.

Natural Language

Unit of "word" is not difficult to treat and to understand for various usages. But "sentence" cannot be treated like "word" as mentioned. And what is more, if it is treated in a context, we have much difficulty in some cases. So an ultimate purpose of natural language study is to learn "understanding of any sentence." It is said that studies of natural language in understanding has three categories: syntactic, semantic, and pragmatic. Syntactic is to structure and to make the grammatical relationships between words in sentences. Semantics is termed with assigning meaning within various syntactic contexts. Pragmatic is to relate individual sentences one another in the whole context category. So the problems of (HC), (NA), and (IH) are still left as the difficulties.

Computer Vision

The purpose of computer vision is to find and to understand a target object in the surroundings. This is initiated in a production spot in order to find a target object and to treat (examine, grasp, fixing, etc.) purposely. And this study comes further to understand an environment by recognizing objects by edging of the objects individually in order to recognize an individual object. And this tendency goes beyond further up to recognition of environment itself including road, tree, moving objects, etc.

So the problems of (HC), (NA), and (IH) are still left as the difficulties.

Expert System

Expert systems is initiated by an idea of imitation a professional expertise behaviour of management and production. A professional behaviour of expertise is ruled out of examining their behaviour one by one in detail. And then those behaviour are programmed in a computer in order to use by anyone needed. But to rule out of behaviour is not an easy task to do at all and the core problem is to understand a professional expertise behaviour by another. So the problems of (HC), (AN), and (IH) are still left as the difficulties.

Theorem Proving and Logic Programming

Theorem proving refers to the process of making logical deductions starting from a non-contradictory set of axioms specified in the predicate calculus. Robinson (Robinson 1995) showed how it was possible to totally automate this process by using a method called resolution. The idea and principle of resolution method is based upon theorem-proving research. And he says further on: "Any assertion that is to be proved using resolution theorem-proving techniques is first represented as a formula in the predicate calculus and then its negation is added to the set of axioms. A purely mechanical set of transformations can then be carried out by the theorem-proving program to put the axioms and negated assertion into so-called 'clause' form. Any pair of these clauses can be 'resolved' against one another in a matching process that results in the creation of a third clause that logically follows from the previous two clause. The resolution method automatically performs a series of such resolutions, eventually building a tree of clause, resolute, resolute of resolute, and so on. If eventually two clauses resolve to 'nil', then the two clauses are contradictory which, in turn, implies that the initial set of clauses is contradictory and this implies, finally, that the assertion has been proved since its negation caused a non-contradictory set of clauses to become contradictory." So the problems of (HC), (NA), and (IH) are still left as the difficulties.

Knowledge Representation

For a system to be able to behave intelligently it must have knowledge of expertise as mentioned in Expert System in the above. In a sense, Knowledge Representation is a process before ruling out of behaviour in terms of logical steps that are related with semantic networks by nodes and combinations. One of a common nodes and combinations in semantic networks is "ISA" link which is able to make facts to be attached to classes of objects and then inherited by specific objects in the class.

So the problems of (HC), (AN), and (IH) are still left as the difficulties.

Learning

At first, learning had not been a major concern for Artificial Intelligence and Knowledge Base. Most Artificial Intelligence researchers seemed to even ignore this kind of study. But after an idea of autonomous behaviour in artificial intelligence came up to as a current issue, leaning has begun to take on ever increasing significance in Artificial Intelligence. So the problems of (HC), (NA), and (IH) are still left as the difficulties.

Dynamic Knowledge Base

Simply to make an artifact like a certain kind of human knowledge which is able to understand a matter, a knowledge, and a figure is being done with various methods. For example Sugiyama (Sugiyama 2004, 2006, 2008) is introducing, the ideas that some kinds of human knowledge are able to be represented and be stored into a computer by using the back propagation neural networks (BPNN). And he proved that these kinds of the knowledge include "Remembering," "Retrieving," "Deducing," and some kind of logical "Thinking-out." However, the abilities achieved by those methods are quite different from the human knowledge that is able to Think-Out dynamically. But this is the main and the most important issue for getting the mechanisms of Artificial Intelligence by using a computer. So this will be the one of the key field in Artificial Intelligence studies.

However, the study itself is under way and no direction for study has been even acquired yet. So the problems of (HC), (NA), and (IH) are still left as the difficulties.

Distributed Computing Method

Distributed computing creates the foundation for manufacturing mediated via intranets of the Internet. In its simplest form, distributed computing having tens and thousands or more computers connected by a network that is working to solve the same problem. Real manufacturing is more complicated, that is to say, machines at a factory have got each computer and the machines should be coordinated/collaborated. The coordination collaboration pattern is not unique but it depends on the type of manufacturing and the company. So usually it is hard to manage the whole system. However, the Internet and intranets are used for getting them work together.

So the problems of (AP), (HC), and (IH) are left as the difficulties.

Agent Method

The Agents are intelligent knowledge bases that can be connected through intranets or the Internet to gather information on purchasing, designing, planning, production, marketing, planning, etc. But it is quite hard to make a proper agent that is able to satisfy the demands and needs because an agent is expected to behave like human. And it seemed that there would not have clear picture and clear definition to an agent logically.

So the problems of (AP), (HC), and (IH) are left as the difficulties.

Holonic Method

The word "HOLON" is firstly used in the book called "The Ghost in the Machine" by Auther Keostler. And he introduced the idea of "Self-regulating Open Hierarchic Order (SOHO)." This can be expressed as "YANUS."

And it says that SOHO has characteristics as;

1. The same as the living creature, it consists of sets of parts and does not consist of uni-motivated chains.
2. The whole of each creatures is branched into sub-wholes one after another autonomously and itself has a hierarchy of multi-levels.
3. It has Self-regulating Open Hierarchic Order; YANUS.
4. It has an autonomy and an integrality.
5. It has hierarchy and networking.
6. It has regulation and it is targeted.
7. The hierarchy has its own ordering regulations.

The studies have started first in Europe, and then Australia, then USA, and in Japan. HOLON is defined in the Consortium in Europe as follows.

HOLON: An autonomous and co-operative building block of a manufacturing system for transforming, storing and/or validating information and physical objects. The HOLON consists of an information processing part and often a physical processing part. HOLON can be a part of another HOLON.

Autonomy: the capability of an entity to create and control the execution of its own plans and/or strategies.

Co-operation: A process whereby a set of entities develops mutually acceptable plans and executes these plans.

Holarchy: A system of HOLONs that can co-operate to achieve a goal or objectives. The holarchy defines the basic rules for co-operation of the HOLONs and thereby limits their autonomy.

Holonic Manufacturing System (HMS): a holarchy that integrates the entire range of manufacturing activities from order booking through design, production, and marketing to realize an agile manufacturing enterprise.

Holonic attributes: attributes of an entity that make it HOLON. The minimum set is autonomy and co-operativeness.

In Europe this is still under the development and study for the aim of application in industrial companies. They have a prototype model of idea of HOLON, but it is that only the idea of HOLON is taken philosophically and the model behaves holonically under only a few conditions. That is today, it does not have a mathematical expression at all. In Australia, US, and Japan, the situation is almost the same as Europe, which means that the philosophical ideas have been implemented in systems, and there have not been introduced usual way of control methods.

By using the idea of HOLON, any complex system or organism is able to be expressed and so this idea can be applied to an engineering control field for a huge and very complex system. But this idea is still in need of a mathematical investigation and study in order to treat it logically and systematically. And the problems of (AP), (HC), (NA), and (IH) are still left for refinement and for constructing a basic idea for real usages.

Chaos; Complexity Theory

"Chaos" by James Gleick became a best seller in 1987. Since then, this world seemed to be caught up by this idea. And many kinds of the phenomenon in this world have been explained or expressed or described by this idea. Since then, mathematical explanations have been given to this idea, but it is still under-going theme to solve as Stuart Daw who is a researcher at the Oak Ridge National Laboratory in Oak Ridge said at the fifth annual conference on "Chaos in Manufacturing" in Santa Fe, N.M., 1997.

"No one can yet define complexity or say exactly how to manage a company in accordance with its principles. But computer tools exist to predict explosions and metal fatigue where once the timing of such a failure was anybody's guess...."

Now we have got the general algorithm (the typical steps) in building a simulation model of Chaos.

Simplify the problem as much as possible while keeping what is essential.

1. The program which simulates many components following simple rules with specified interactions and randomizing elements.
2. Run program many times with different random number seeds, collecting data and statistics from the different runs.Attempt to understand how the simple rules gave rise to the observed behaviour.
3. Perform parameter changes and "lesions" on the program to locate the sources of behaviour and the effects of different parameters.
4. Simplify the simulation even further if possible, or add additional elements that were found to be a necessary.

But this idea is still in need of a mathematical investigation and study in order to treat it logically and systematically. And the problems of (AP), (HC), (NA), and (IH) are still left for refinement and for constructing a basic idea for real usages.

As seeing from the above discussions on an expert system, an agent method, a knowledge base, and an intelligence system in a management system, and as we can easily notice from the focused problems in the above discussions, it is quite important thing to use "Knowledge" for solving the problems (AP, HC, NA, and IH).

MAIN FOCUS OF THE CHAPTER

Here discusses on the main issue of the problems' (AP, HC, NA, and IH) solving by giving some definitions about Knowledge, Knowledge Base, and Artificial Intelligence (Sugiyama 2004e, 2005g, 2006h, 2008i, 2009k). Especially the discussion will be focused on the dynamic behaviour (Dynamics in Knowledge) in processing of Knowledge by using a neural networking (Sugiyama 2008j, 2009k) in order to solve the problem facing.

Firstly here discusses on Artificial Intelligence, Knowledge, and Knowledge Base in General.

Even though in the above Sections there explained and discussed on Artificial Intelligence, Knowledge and Knowledge Base, they are very ambiguous "words" to explain and to define in detail. It may be related with abilities for actions; knowing, understanding, thinking, familiarity gained by experience, and doing things, or it may be related with a behaviour for actions. And also Artificial Intelligence, Knowledge, and Knowledge Base have a very ambiguous "structures" to explain and to define in detail although there are some explanations for those. These facts will show that the present techniques can only be applied to an individual target in certain limited areas within nonflexible technical and system limitations although those system and technique sometimes work such and such individually in very limited usages.

So at present, we may conclude to say that we do not know yet and do not have yet a clear picture of it although it has been challenged to make Artificial Intelligence, Knowledge, or Knowledge Base by using a computer for many years.

So, it is very important and necessary to know more about what Artificial Intelligence, Knowledge, and Knowledge Base are meant to mean. And also it is quite important to find a description method of Artificial Intelligence, Knowledge, and Knowledge Base as a system.

Core Factors and Definitions of Dynamics in Knowledge

As we can easily notice from the above discussions, it is quite important to use Artificial Intelligence, Knowledge, and Knowledge Base for solving the problem (AP, HC, NA, and IH). Because the direction of the behaviour for those is to behave like human. So here further discusses on this key matter in detail.

Artificial Intelligence, Knowledge, and Knowledge Base

All the knowledge of human intelligence has not been fully studied yet, but those have been rationally applied in many fields to bring about an intelligent artificial creature for solving a problem. This clearly shows that the researchers know the limit of an artificial intelligence as a narrow sense of intelligence. And there is small doubt among some researchers that an artificial intelligence will be capable of creating an intelligent thought even in the near future.

At this stage we can say that there are many different approaches to Artificial Intelligence, Knowledge, and Knowledge Base, and that none of which are either completely right or wrong. Some are obviously more suited than the others in some cases, but the others do not.

In the past, Artificial Intelligence research has mostly been focused on solving specific problems. And this phenomenon makes the field of Artificial Intelligence split into many branches; from a pattern recognition to an artificial Life, including Evolutionary Computation, Soft Computing, Fuzzy Logic, and Neural Networking.

Here, we look at Artificial Intelligence, Knowledge, and Knowledge Base from a point of meanings.

A dictionary briefly says about Knowledge/Intelligence as follows.

1. Understanding.
2. Familiarity gained by experience.
3. Within a range of information.

So, in another words, we can interpret the above expression such that “Knowledge” or “Artificial Intelligence” means to understand what is going to do, is doing, and has done for a purpose to be accomplished in a range of information.

However, this is a knowledge statement for Humankind, we need to restate and to redefine for Artificial Intelligence, Knowledge, and Knowledge Base as follows.

“Artificial Intelligence, Knowledge, or Knowledge Base is a mechanism for accomplishing a desired purpose to be done by getting a necessary information one by one from an environment facing through the sensors or by communicating among entities (Artificial Intelligence, Knowledge, Knowledge Base, Data, Objects, etc.) within a range of information.”

At this stage, we can easily understand about the Artificial Intelligence, the Knowledge and the Knowledge Base as the same kind of entity because of the behaviour. So from now on, Artificial Intelligence, Knowledge and Knowledge Base will be written just as “Knowledge.”

Next, we will look at this matter from a point of view in “structure.”

Generally speaking, as it is mentioned in the above, we have a problem in a structural uncertainty. And they are as follows.

1. Not easy to describe and to control a whole system (Environmental Uncertainty; (EU)).
2. Many pre-processes are needed before inputting data onto Knowledge (Data Uncertainty; (DU)).
3. Not easy to understand a rough environment (Functional Uncertainty; (FU)).
4. Poor for a system to understand what Knowledge is doing (Knowledge Uncertainty; (KU)).
5. Not easy to communicate among agents (Communicational Uncertainty; (CU)).
6. No way to re-examine an output quality (Self Recognition Uncertainty; (SU)).

If we look at the uncertainties above (EU, DU, FU, KU, CU, and SU) more closely, they will be focused onto a rigidity of Knowledge. Because we are able to have a nice Knowledge that will behave nicely for solving a problem only if we have a usage within a very limited condition with a very stable environment. That is to say, on the contrary, “Dynamics” in Knowledge is the key matter in order to let Knowledge work appropriately. And so, here gives some discussions and studies on Dynamic Behaviour in Knowledge.

Piaget (Piaget 1999) has given us two views about ‘Knowledge Acquisition’ from a psychological stand point of view. And they are as follows.

Assimilation → “Passive Modification”

1. The modification of an incoming stimulus by the activity of a pre-existent mental structure. Such structures are denoted ‘schemes’, and represent the regular structure of an action.

Example. The general method involved in catching a ball:

The ball catching scheme, producing ‘the catching action’, assimilates the unique circumstances in which the person and the ball may start moving within a very limited behaviour.

Accommodation → “Active Modification”

2. The active modification of the structure – scheme – itself, so as to adapt to the situation.

Example. Catching the ball from different directions:

An initial attempt is made to assimilate the object into a general ball catching scheme. If this fails – the object cannot be caught by the method of assimilation – then an accommodation takes place. The catching scheme is adapted, for instance, by adjusting an arm form by twisting the arm for the glove to face the ball, and then re-applied and happily could get the ball. It may be retried again if it does not go well by collecting another scheme of catching the ball.

So from the facts about the human knowledge acquisition mechanisms from Piaget, we can conclude to have the following views about Knowledge.

1. Passive Modification would cause a problem when Knowledge faces an unexpected situation.
2. Active Modification would be better off in behaviour for Knowledge to overcome an unexpected situation.

Here looks at the case of living creatures what will happen for problem solving as behaviour in action. The following shows about functions of brain what to do in action. In general, a living creature's brain will be functioned as follows.

1. At first stage, there is stimulation from the world to the five senses of a living creature, or from an internal brain itself.
2. Next, capturing the stimulation through the senses or by the frontal cortex. The stimulation will be transformed into 'a semantic form' in a sense of words or a real image or an abstraction in order to understand what it will be.
3. Taking an action according to this understanding.
4. If the input (the requirement) is satisfied by the action, it completes the process and ceases to activate the group of neurons.
5. If it is not satisfied, taking a next action for the input accomplishment by introducing a next related group of neurons or a next related knowledge within the numbers of co-related group of knowledge. This process will be able to go on and on until it reaches to some limitation that makes the process cease to act as "Active Modification" in a case of satisfaction to the input, or in a case of the nature of neurons.

In the above events, there are several events ((2), (3), (4), (5)) that are not good at processing for the present "Knowledge" system.

From the weak processing events shown above, we can extract Core Properties existed to show below.

A. A General Expression in Description of Knowledge.
    1. System - Knowledge - Action - Event - Core Factors - Group of Processing.
B. Forming Co-related Group of Knowledge with a semantic form.
C. A method of Active Modification within the co-related groups.

Solutions

In the above section, we look at the core factors of "Active Modification" in Knowledge, which will assist and solve a problem facing.

Here shows the core factors for solving the problems [(AP, HC, NA, and IH), (EU, DU, FU, KU, CU), (a, b, c)], and SU by introducing "Knowledge" as "Notion."

Notion 1: (Primitive Element)

As we know about human brain mechanism, input information to eye, ear, etc. are broken (transformed) into pieces at one mediator to another in brain in order to get very primitive elements of those. With those very primitive transformed elements, human brains can recognize and understand things from outer world.

We can use this mechanism of human brain (primitive element) in order to make a system communicate with others (entities).

Notion 2: (Hierarchy of Multi-levels)

Hierarchy of multi-levels (a finite set of whole U, finite sets of parts {∑Pi} and finites sets of subparts {∑SP1j},.., finites sets of sub-subparts,.. {∑SPnk}) of system of organization) is defined mathematically. And the structure of the system's elements' connection at each layer is defined by {U⊇∑Pi⊇∑SP1j⊇...⊇∑SPnk}. And these U, Pi, SP1j,..., SPnk are correspond to each layer; Layer-1, Layer-,2 Layer-3,.., Layer-(n+2) in the system of organization, where i∈[0, m0], j∈[0, m1], k∈[0, m2], n∈[0, m3], m0, m1, m2, m3 are positive integers. Hierarchy of multi-levels is made arbitrarily by grouping in terms of a function's mechanism, a function's attribute, and a function's physical figure together.

We can use this mechanism of Hierarchy of Multi-levels in order to make a system transformed into very primitive stages.

Notion 3: (Networking, Regulation, Targeted)

The system of organization that has Networking, Regulation, and Targeted factors of being themselves networked with one another in a regulated way. For targeted has a characteristic of independent itself as well as dependent to another.

We can use these functions of Networking, Regulation, and Targeted in order to make a system combine one another in elements of all over the system and multi-levels.

Notion 4: (SOHO)

Self-regulating Open Hierarchic Order (SOHO) is defined by a property that each of U, Pi, SP1j,..., or SPnk has its own independent system of organization. And those of U, Pi, SP1j,..., SPnk are independent themselves and also dependent one another as well. And each of those is self-regulating.

We can use this mechanism of SOHO in order to make a system flexible to changes.

Notion 5: (Interchange of Whole and Part)

Every factor of U, Pi, SP1j,..., or SPnk in a system has a moment to be the whole of it and at another instance the whole may become a part of it. This topological interchange of the whole and a part can occur when U∩Pi∩SP1j∩...∩SPnk =Φ. When the interchange between the whole {U} old and an arbitrarily Ψ⊆{U}old is made, the topological transformation from the old hierarchy {U⊇∑Pi⊇∑SP1j⊇...⊇∑SPnk}old to the new hierarchy {Unew⊇{Ψ, ∑Pi∩Ψ}new occurs.

We can use this function of Interchange in order to make a system simple and targeted.

Notion 6: (Mathematical Expression for Knowledge)

Let X denote the set of arbitrary primitive data through the preprocessor from the outside world and assume there are (m x n) real valued functions. X is mapped into R, which is the set of relevant knowledge. It is defined as below.

Φij(x):X→R, i=1,...,m,

Ωj(r):X→R, j=1,...,n.

Each element of X is producing a primitive data through the preprocessor for each pattern and each element of R is producing an output to X.

Writing Φj(x) to denote the value of the ith primitive data of pattern x, the vector Φj(x) is shown as follows,

Φj(x)=[Φ1j(x),.., Φmj(x)]T.

Writing Ω(r) to denote the value of jth knowledge base of entity r related with Φj(x), vector Ω(r) is shown as follows,

$\Omega(r) = [\Omega 1(r),.., \Omega n(r)]T$.

The above are feature vectors description of x and r.

There are no restrictions on the forms of the functions Φij; they can be linear or non-linear functions of X.

This makes possible to express the above Notions mathematically and logically in order to treat Knowledge as an entity that will be implemented in a computer system.

Notion 7: (Select Any Part)

Let the whole universe {U} consists of

$\Psi = \{\Psi 1,.., \Psi n\}$

as shown below.

$U \supseteq \Psi$.

Ω(r) is correspond to Ψ each other as shown below.

$\Omega 1(r) \rightarrow \Psi 1$,

$\Omega n(r) \rightarrow \Psi n$.

The Reflex Transformation is defined as,

$U \rightarrow \Psi k \supseteq \{U \cap \Psi \cap \Psi k = \varphi\}$.

This also makes possible to express the above Notions mathematically and logically in order to treat Knowledge as an entity that will be implemented in a computer system.

Notion 8: (Dynamics in Knowledge)

It was possible to organize "Knowledge" in a conventional method within a very limited behaviour and actions with very much rigidity by using Back Propagation Neural Networking (BPNN). And by adding the ideas of "Dynamics in Knowledge" (Sugiyama 1994a, 2001b, 2002d, 2004e, 2005g), we can make "Knowledge" that will behave dynamically in order to select a related knowledge out of the related group of Knowledge.

This is shown by using a simple example by using BPNN as shown below.

Firstly, "Knowledge" in conventional sense is shown. That is to say, whenever a complex Knowledge is made, the behaviour is simply to produce the output even though the output is not an expected one as illustrated in Figure 1.

This Knowledge does not behave properly when unexpected event occurs. As mentioned in the previous sections, the reason is that this Knowledge is lack of flexibility in change.

But the Knowledge shown in Figure 2 below has a flexibility to select desired and related group of Knowledge in a category by Linker from the content related database of accumulated Knowledge, which is able to search and to form a related kinds Knowledge as a group. The group is selected by the value of Co-Relational Factor, which is given by BPNN output value.

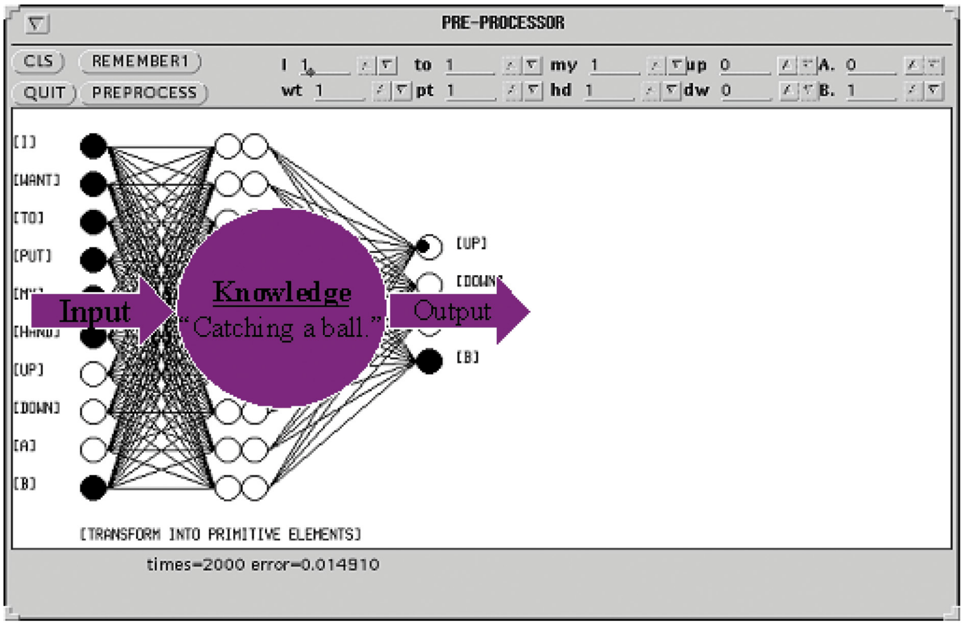
By using the group of related Knowledge, a system can select the most appropriate Knowledge from time to time. That is to say, if there is a problem that goes good in a certain situation, the system tries to find another most appropriate Knowledge for this situation as shown in Figure 3.

This example showed that this Knowledge could behave like in a case of living creatures what will happen for problem solving as in action. That is to say, the system is able to select a proper Knowledge from a group of the related Knowledge in order to face to a problem solving.

Notion 9: (General Behaviour)

By using these Notions, "Knowledge" is able to have a hierarchical structure and is also able to reconstruct for targeted, so that a necessary knowledge is able to be given one after another for

Figure 1. Conventional knowledge presentation in general



re-processing in order to get an improved output through communication.

This says that any necessary knowledge Ψk can be extracted, can be targeted, and can be concentrated in order to get the desired output.

So by using Notions 1 to 9 stated above, we can expand any system (system, agent, organization, etc.) hierarchically with multi-levels of layers, linked and networked. And also each element at each layer can easily communicate with another because of its primitively hierarchical structure.

Figure 2. A group of related knowledge

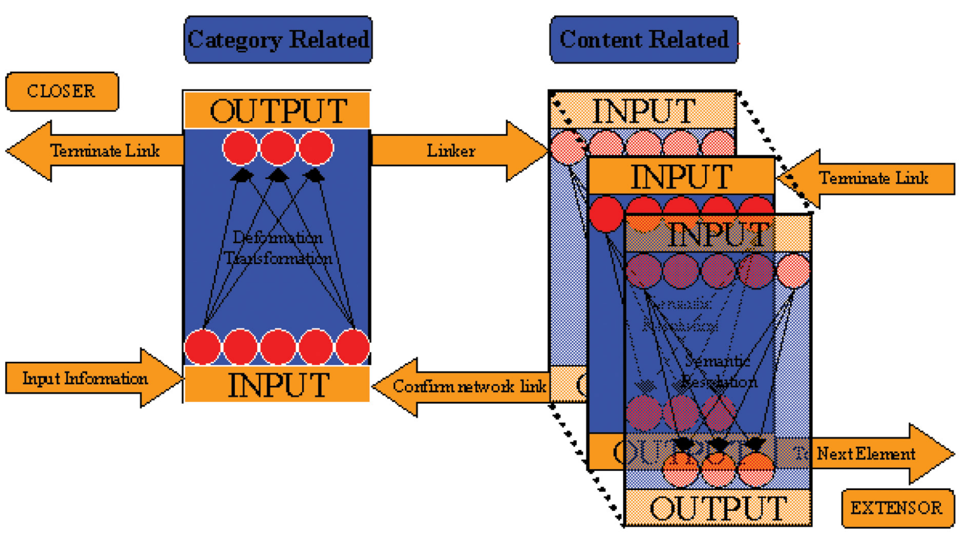
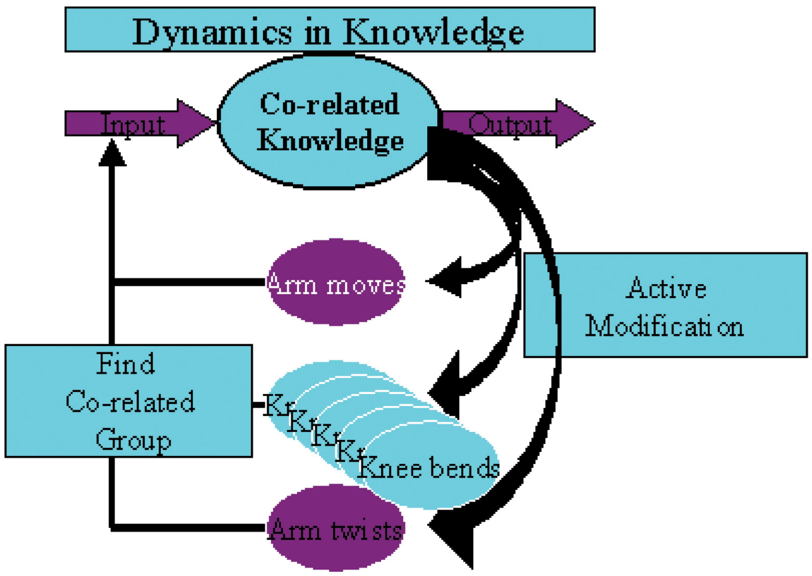


Figure 3. Active modification for new knowledge



FUTURE RESEARCH DIRECTIONS

Here shows issues left for studies on Dynamics in Knowledge in order to get further developments.

A flexibility of Knowledge is given by the ideas shown above sections. But it goes better in the behaviour of Knowledge if Knowledge can communicate each other like human. The Knowledge studied here consists of primitive elements by BPNN, so that some kind of communication between Knowledge is capable to make but it is quite a primitive level of communication in the behaviour. One tip for solving this issue is to make a general Knowledge for each category that will understand a whole context of each category.

The flexibility for each primitive Knowledge (atomic Knowledge) in a sense of structure is also important matter. That is to say, each Knowledge that will be able to change its structure as describe in the Notion 5 is also very important matter to have. Because it is quite easy to get a proper output if Knowledge is transformed into the figure of Knowledge that is most suitable one. For example, if there is the Knowledge of searching an object, it would be better to transform the figure for a particular object. Because "Searching a pen" and "Searching the meaning of Pen" is different from each other.

And it is strongly expected to study on data structure for Knowledge. That is to say, whether or not it would be better to have data for each Knowledge or for the whole Knowledge or to be unified?

CONCLUSION

Through the examinations and discussions in this paper on Knowledge related matters and Knowledge itself, the following results have been given as shown below by touching upon a basic level of control, management, knowledge base, and artificial intelligence.

1. Basic idea of the Hierarchical Multiple Knowledge method has been introduced.
2. General mechanism of dynamic behavior of Knowledge through communication that behaves more intelligently than before has been introduced.

3. Overall system description method in complexity and hugeness by using the idea of Hierarchical Multiple Knowledge has been introduced.
4. A new direction and a method in "Knowledge" is logically expressed by inaugurating the brad new idea of "Dynamics in Knowledge," which will behave more flexibly and intelligently in real usages.

REFERENCES

Sugiyama, S. (1994a). *Semantic and logic theories in neuron networks*. IEEE International Conference on system, Man and Cybernetics.

Sugiyama, S. (2001b). *Reflex transformation of consciousness*: *Toward a science of consciousness*.

Sugiyama, S. (2002c). *Basic model for muscle control in primitive neural sets with reflection*. First World Congress on Biomimetics. CDROM version.

Sugiyama, S. (Ed.). Imai, S., Kobayashi, T., Takeuchi, A., Hirayu, H., Miwa, S. (2002d). *Investigational research report: Studies on next generation industry and company system.* Tokai Techno Highland.

Sugiyama, S. (2004e). *Augmented agent for multi-purposes*. World Congress Biomimetics, Artificial Muscles & Nano-Bio 2004.

Sugiyama, S. (2004f). *Agent talk in neural networks*. ANNIE.

Sugiyama, S. (2005g). Multiple layer control method. In J. A. Meech & Y. Yoshiyuki Kawazoe (Eds.), *Intelligence in a Small Materials World: Selected Papers from LPMM-2003*, (pp. 286-298). ISBN1932078193

Sugiyama, S. (2006h). *Tele diagnostic by Web*. SPIE2006 Medical Imaging, CA.

Sugiyama, S. (2008i). *Fundamental behaviour in communication method*. IEEE/INFORMS International Conference on Service Operations and Logistics, and Informatics (IEEE/SOLI'2008).

Sugiyama, S. (2008j). *Ubiquitous framework in service science. The 2008 Logic and Science of Service*. The New Wealth and Wellbeing of Nations.

Sugiyama, S. (2009k). *Feature extraction in system*. INFORMS International Conference on Service Science (ICSS'2009). Hong Kong.

Sugiyama, S., & Burgess, L. (2011). Advanced design approaches to emerging software system: Principles, methodology, and tools In Liu, X., & Li, Y. (Eds.), *Principle for engineering service based system by swirl computing*. Hershey, PA: IGI Global.

ADDITIONAL READING

Anderson, J. A. (1989). *Introduction to neural networks*. Cambridge, MA: MIT Press.

Arbib, M. A. (1989). *The brain as network of neurons. The metaphorical brain*. New York, NY: Wiley-Interscience.

Barth, D. (2004). *Mammalian neurophysiology: Sensorimotor system I*.

Bertalanffy, L. (1950). The theory of open systems in physics and biology. *Science*. (n.d)., 111.

Bloom, C. A., & Lazerson, A. (2008). *Moving brain mind and behavior*.

Bloom, F. E., Lazerson, A., & Hofstadter, L. (1985). *Brain, Mind, and Behavior*. Educational Broadcasting Corp.

Elman, J. L. (1993). *Learning and development in neural networks: The importance of stating small, cognition*.

Piaget, J. (1999). *The MIT encyclopedia of the cognitive sciences*. MIT.

Robinson, J. A. (1995). *A machine-oriented logic based on the resolution principle* (pp. 23–43). ACM.

This work was previously published in Knowledge Management and Drivers of Innovation in Services Industries, edited by Patricia Ordoñez de Pablos and Miltiadis D. Lytras, pp. 102-119, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 86
Tool and Information Centric Design Process Modeling:
Three Case Studies

William Stuart Miller
Clemson University, USA

Joshua D. Summers
Clemson University, USA

ABSTRACT

A new design process modeling approach focused on the information flow through design tools is discussed in this chapter. This approach is applied to three long term mechanical engineering design projects spanning 24 months, 12 months, and 4 months. These projects are used to explore the development of the new modeling approach. This is a first step in a broader effort in 1) modeling of design processes, 2) establishing case study research as a formal approach to design research, and 3) developing new design process tools. The ability of engineers to understand the dynamic nature of information throughout the design processes is critical to their ability to complete these tasks. Such understanding promotes learning and further exploration of the design process allowing the improvement of process models, the establishment of new research approaches, and the development of new tools. Thus, enhancing this understanding is the goal of this research effort.

INTRODUCTION

This chapter begins with a general discussion on the design process and approaches to capturing and modeling the design activities. It is not intended to be a comprehensive review of the literature, but to provide a frame of reference with respect to the critical issues associated with design processes. This is followed by a discussion of case study research in engineering design. Again, this section is used to provide the reader with a broad understanding of how current case study research has been undertaken in engineering design and how it can contribute to our understanding of the

DOI: 10.4018/978-1-4666-1945-6.ch086

design process. Next, the reader is introduced to the relevant issues associated with representing the information flow through the tools used throughout the design process. Finally, three industry focused case studies are used to illustrate the application of this tool, demonstrating how it can be used to highlight issues such as iteration, information dead-ends, and tool function duality.

THE DESIGN PROCESS

The design process is a flexible, high level, logical network of activities to be performed and/or design tools to be used for the entire act of designing an artifact, formed by choosing desirable candidate(s) from a set of viable activities/design tools based on certain objectives (Hazelrigg, 1998). The design process is the collaboration of scientific "know how" with mental and physical steps being taken toward the goal of arriving at a satisfying solution (Simon, Kotovsky, & Cagan, 2001). It is a social activity which allows the generation of physical and intellectual property from mental organization and physical tasks (Leifer & Tang, 1988). Engineers perform design processes often with varying degrees of success. The goal of the research presented here is to enhance the ability of designers to understand and therefore complete design processes.

Design Process Modeling

Several approaches have been proposed in the literature for modeling design processes and activities. This section is not intended to review exhaustively the different approaches to modeling engineering design processes, but to highlight a fundamental limitation that appears to be common in most approaches. This limitation is their inability to easily capture the activities executed by the engineering designers in larger design projects outside of artificial academic environments. The goal of this research is to overcome this limitation by focusing on the actual engineering design tools that are used and to model the flow of information through these tools. This modeling approach will be discussed in the following sections.

One example of a design process modeling scheme is to view engineering design as a series of transformative steps that can be modeled as state transitions (Reymen, et al., 2006). In this approach, each state of the design, captured as the complete collection of the description of the design product, is modeled with a transition activity linking the initial and final state. A limitation of this approach is that the granularity of defining the transition activities is extremely coarse and there is no clear demarcation between states, as these are an artifact of the process modeler, rather than the design process itself.

Ullman presents a different approach to modeling the design process based on his empirical studies in the early 1980's (Ullman, Dietterich, & Stauffer, 1988). This approach seeks to define the task episode accumulation (TEA) model as a sequence of tasks that are essential information processing activities. The different activities are classified to include such actions as gather, synthesize, and decide. In this manner, the TEA model looks at sequencing the actual actions taken by individual designers as opposed to capturing the transition states between the entire collection of design product information. This model has been used to capture designer thought processes in controlled environments, but not in larger design projects as it would quickly become too unwieldy to process each task.

Another transformation based modeling approach to engineering design is that of the Function-Behavior-Structure model (Gero, 1990). In this model, there are eight basic steps that transform a function objective into a structure to realize that objective. While this model suggests that there is a distinction between different types of information and defines different design activities as formulation, synthesis, analysis, evaluation, documentation, and three types of reformulation, this model does not adequately capture all of the details of the design activities and it does not include other design information types, such as requirements.

Requirements tracing and design rationale capture is the basis for the design process model that has been proposed by (Burge & Brown, 2002). Essentially, this model creates a decision tree where design solution options are explored. At each branch, a criterion for success is selected and a branch is chosen. The rationale for selecting the option branch is defined. These three steps form a design task that is sequenced together to form a design plan. These tasks are used to define the value of the specified engineering design variables, thus linking the design requirements with the final design product. This model of engineering design provides for explanations as to why and how decisions are made. However, in large design projects, this level of documentation is not typically available thereby causing gaps in the modeling of the complete design process.

A different approach to modeling engineering design processes is to view the entire "space" of the possible design activities that can be used to achieve the artifact (Grabowski, Lossack, & Weis, 1995). In this model, the order of the design decisions leads to a path through this process space to the concretion of a design artifact. The activities, or directions, that can be chosen to explore this space include: concretion, abstraction, detailing, combining, varying, and limiting. These are similar to the activities defined by Ullman. This model provides, perhaps, an interesting theoretical view of the design process, it is not clear how this model could be constructed to capture a large scale design project.

Finally, Ostergaard and Summers present a model of engineering design from the perspective of the flow of engineering information through a series of design resistances that are the collaborative design situations (Ostergaard & Summers, 2007). In this model, the design information is represented as the current and each stage where information is passed between human and computer reasoning agents is treated as a resistance to this flow. This resistance is overcome through the application of external knowledge, termed design voltage. While this approach begins to capture the concept that information is accumulated throughout the design process through transformations, it has proven difficult to quantify and model actual engineering design projects.

These examples of different approaches to modeling the design process have been developed for many different purposes, including understanding how designers think, defining how computational reasoning support tools can be developed, how states of engineering design products evolve, and the flow information through collaborative situations. Unfortunately, none of these approaches have been easily applied to capturing the activities within realistic industrial engineering design projects. A model engineering design activities that can be used for studying large scale design projects.

Metrics of Design Processes

The design community has several measures of design success (Yang, 2007). One of these, product functionality, relates to how well the design solution accomplishes the tasks it has been given (Pahl & Beitz, 1996). Another common metric is customer opinion, a subjective measure. This metric can vary with aesthetics, comfort, appeal, or market trends (Kirschman, Fadel, & Jara-Almonte, 1996). Another more quantified success metric is the designer efficiency (Atkinson, 1999). This measure is used to find the time and cost to design. This could be measured as the time the designers take in the design process to develop the final product. It is recognized that many different methods of success evaluation in design exist (Ulrich & Eppinger, 2008). However, discrepancies exist in the academic community about which success measures are valid (Sobek II, 2007). Ultimately, only a handful of success metrics for design processes exist and there is no consensus which should be used.

All success metrics can be used by a variety of users. Designers, managers, and researchers all use success metrics of processes. Naturally, each user has their preferred success metric which caters best to their particular area of interest. However, each category of the metrics can be evaluated independently and used to highlight the most desirable traits within the options given. An absolute metric would be beneficial over a relative metric because any ambiguity about the results would be eliminated. A direct measure of the design process is desirable because directly measuring the process leaves less room for interpretation errors that are present when the products are evaluated.

For design processes to be comparatively evaluated through different metrics, the processes and their goals must be clearly defined and then related to each other. To fully explain the process, it should be broken down into sections or stages which can give more detailed understanding of the change of information throughout the process. Doing so not only gives designers understanding of what the process accomplishes, but also how it is accomplished. Each stage can be further analyzed to determine the individual steps taken within that stage to achieve the deliverables and how those interact with subsequent stages. Each of these steps can be evaluated to determine the information that enters from the previous step and the information that exits to the next step. The tools used within the design step can be identified as well as their fundamental function.

Decomposition of the Design Process

Design steps can occur in loops, repeating and converging on information vital to success. Each design tool or method receives information from other sources, and transmits its exiting information outward to another step. Steps can possess singular and multiple information units which can enter and exit specific process steps. The information produced by any design step should have value greater than the sum of its components. The same can be said regarding the design process and its subordinate stages. Figure 1 illustrates a design process that has design stages decomposed into design steps. This notion of decomposition of the general design processes into constituent elements has been the focus of research efforts that yield ontologies of design activities at the atomic level (Kumar & Mocko, 2007; Sim & Duffy, 2003; Whitfield, Duffy, Coates, & Hills, 2002). From a pragmatic perspective, engineers and practitioners would likely prefer the granularity of resolution that focuses on design tools rather than atomic cognitive and information manipulative activities. This view through the design tools is the focus of this research.

In order to quantify how well a design process has been executed, the process itself must be analyzed (Lockledge & Salustri, 2001). Analyzing processes can be done with various levels of detail resolution. A low detail resolution of the process allows the researcher to retain focus on the overall process goals. It does not lend detailed insight to information interactions within the process. Figure 2 shows a low resolution approach to process analysis. This type of analysis builds on the establishment of what is entering and exiting a specific boundary to the process. The result is a coarse representation of the process.

Conversely, in Figure 3, a higher detail resolution analysis of processes consuming more time and effort in detailing each step is shown. The work becomes tedious and can go into increasingly deeper detail, which may or may not be needed. This intermediate resolution of detail shows how increasing detail requires increasing effort as well, because the entities being considered increases, thus increasing the work required to analyze.

This type of analysis permits the researcher to identify information interactions within the process, thereby identifying additional information

Figure 1. Design process hierarchy

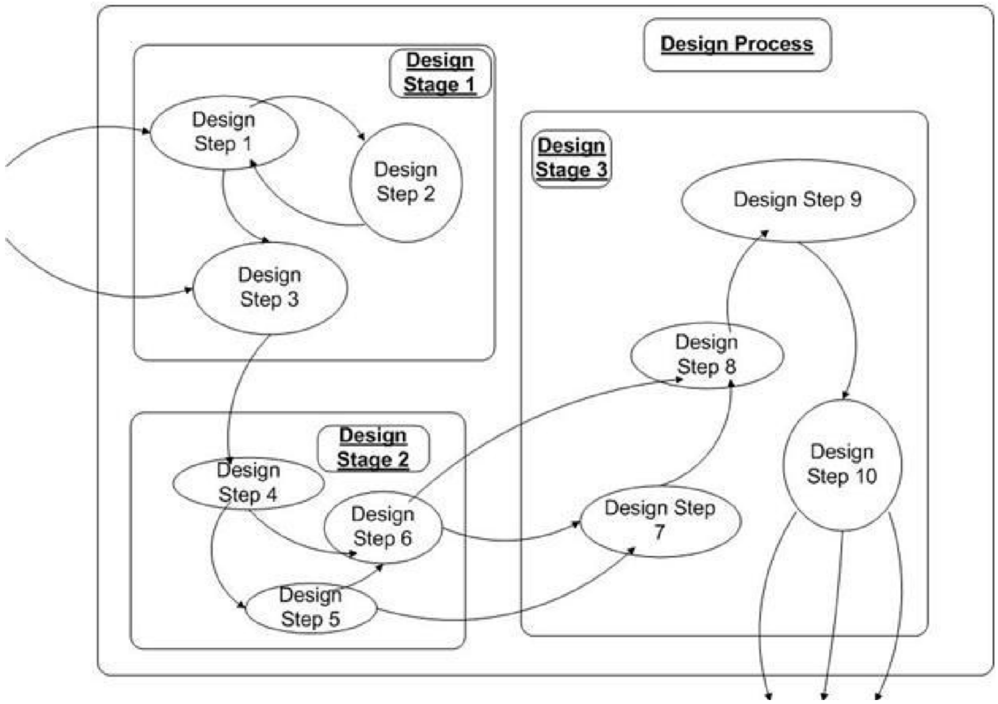


Figure 2. Low detail resolution

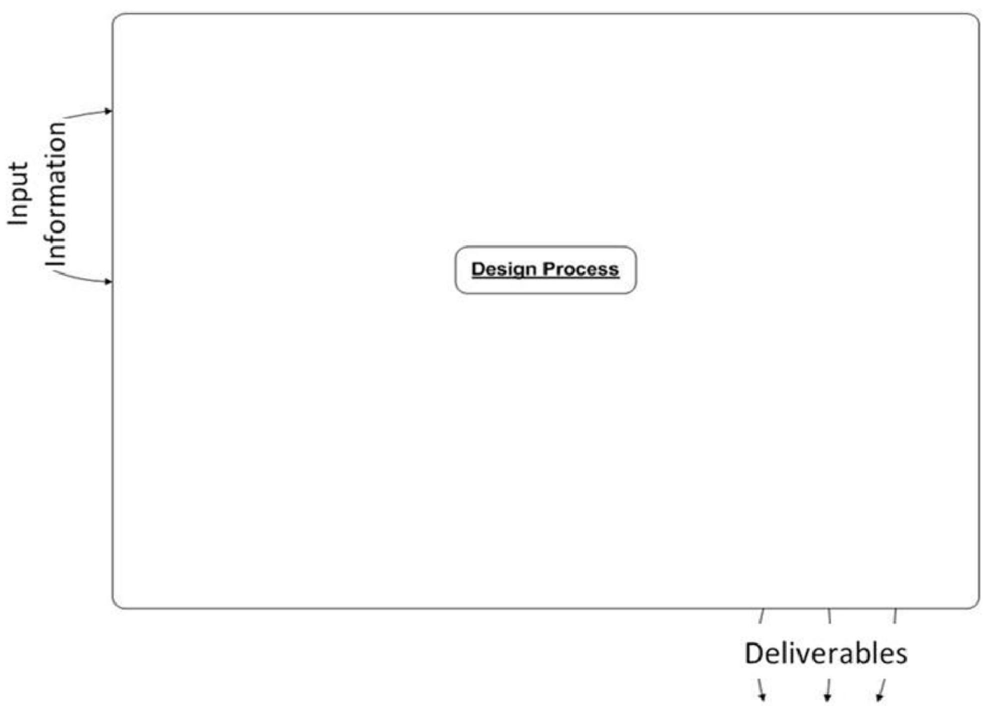


Figure 3. Intermediate detail resolution

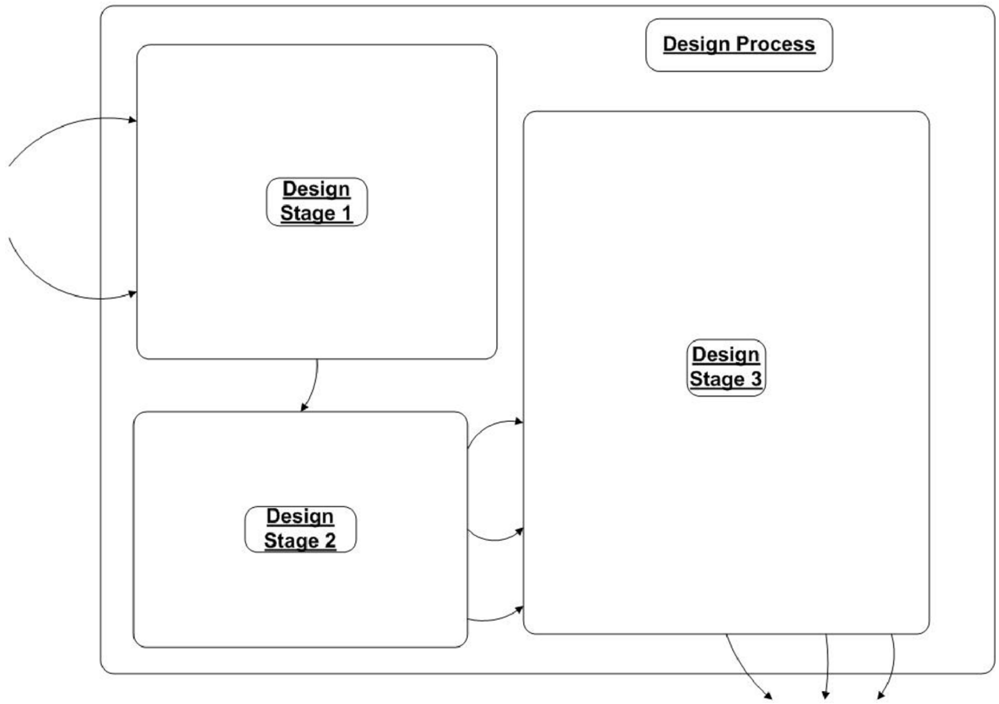
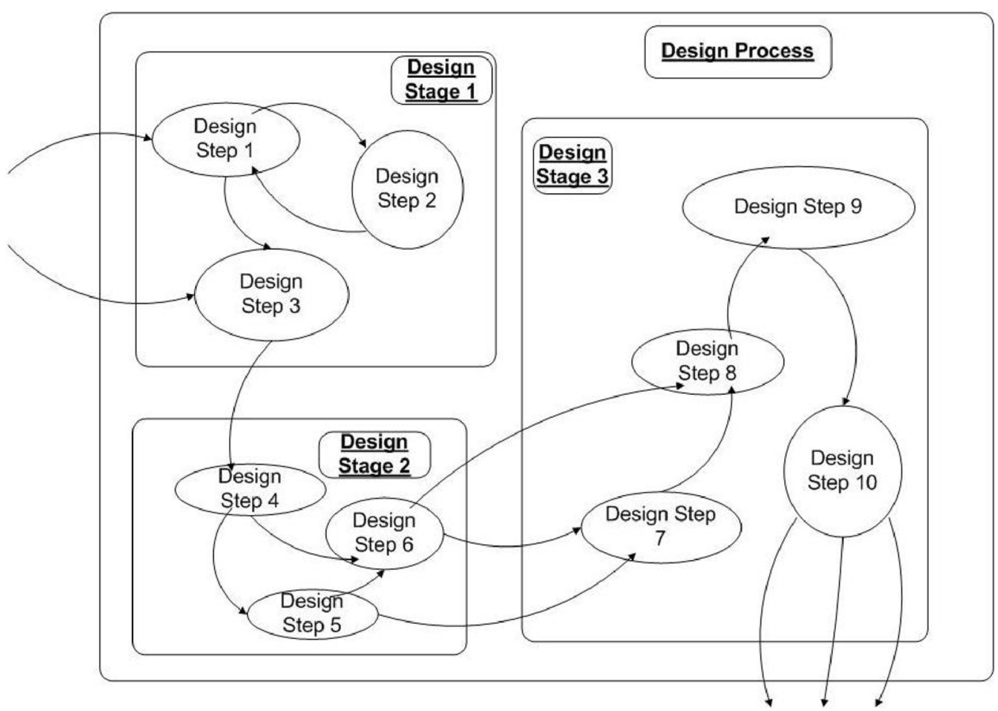


Figure 4. High detail resolution



within the process that was initially overlooked. A high detailed resolution analysis of the design process is shown in Figure 4 which is an extension of the hierarchy defined above. This breakdown of the process is considerably more "incremental" than a low detail resolution analysis can give. It is the designer, or as previously discussed, manager or researcher who must be capable of specifying appropriate detail to consider in the evaluation that will not consume enormous processing power yet still yield a sufficient analysis.

Design Process Tool Support

Design tools can exist in two different forms; product support and process support. Product support tools are used within the design process to complete and achieve the design solution. Process evaluation tools are used to evaluate and observe the design process itself, thus giving some measure of goodness. These tools may or may not be used to arrive at a solution, but they are frequently used to manage and monitor the design process either en route or after a solution has been achieved.

Analysis of design can be accomplished with several existing tools, albeit with limitations. Program Evaluation and Review Technique (PERT) is a design tool which is used in scheduling and planning processes (Battista, Pietrosanti, Tamassia, & Tollis, 1989). PERT is best used to evaluate singular project management entities, such as time. Gantt Charts are another project or process management tool. They are used to present time related information of processes in relevant timeline domains (Maylor, 2001).

The ability of the user to modify the evaluation tool for their specific needs does exist, but is often tedious due to evaluation tool restrictions of formatting and application. A process evaluation and analysis tool is needed to allow the users to track the critical information that is important to their individual research. The ability to decompose processes and evaluate each sub-system should prove beneficial for designers, by increasing their understanding of the design process.

Such a tool should facilitate information tracking and evaluation for each step within the process and should display the information produced by performing the specified tasks within that process. If units of information, such as design documents, are generated by a design task, they must either be useful or considered waste. Wasteful information means that the information is not useful in producing a solution and should not be created. If information is created and used later on, but not part of the original intention of that step, the requirements should be revised to include the newly discovered information. This analysis should enable researchers to better understand design processes. Researchers can analyze design projects to use the information available to learn about the process while using the process to learn about the applied case as well.

CASE STUDY RESEARCH IN ENGINEERING DESIGN

One of the most significant tools for design process analysis is case studies (Yin, 1994). Remarkably, case studies are often misunderstood when used as a design process analysis tool (Ahmed, 2007). Case studies are the empirical extraction of data from real world events that are used to view relationships and examine results about design (Teegavarapu & Summers, 2008). They provide relative, fact-based results to qualitative questions (Eisenhardt, 1989). Case studies focus on real world practices to develop theories and methodologies (Teegavarapu & Summers, 2008). They generate straightforward data from results which apply to a group of design variables, but cannot lend themselves to distinguish the effects of each variable individually.

Collecting data from case studies can laborious and can take place in a variety of manners, but tradeoffs of collection cost, collection effort, collection time, and influence on the results exist (Yin, 1994). The researcher must select what information is needed as well as how to extract it and do so without skewing the data. To extrapolate the conclusions to a theoretical application, properly constrained studies must be established (Yin, 1994). Developing such a bank of studies is difficult and resource consuming because the similarity of multiple cases is subjective and not easily compared.

Using case studies is said to be an all encompassing method which covers the problem definition, hypothesis formation, and collection and analysis of data stages. Case studies enable the designer to answer how and why questions about the specific occurrence. They do not allow the user to control variables, however, which would classify the study as an experiment. Case studies possess tremendous power in analyzing design processes, but wielding such power requires care. Despite the informational gain potential, case studies are easy to execute incorrectly. This reduces validity of the results as well as the beneficial experience gained by the researcher who executed the case study.

In order to enhance the use of case studies within design, one should be able to visualize the design process, without affecting the products of the process. This would allow the researcher to know how to construct the case study without corrupting the data. The visualization, as well as the results generated from it must be easily observed and understood. The information gained from visualizing the design process should not require exhaustive effort to generate or comprehend thus reducing the probability of erroneous case study design and extrapolation. The identification of multiple similar cases should not require exhaustive efforts but should rather be an observation of the two cases. By improving the use of case studies, engineers can focus more on what can be learned rather than doing tasks correctly.

The Need for a Case Study Research Tool

A design process model is needed. If a design process model existed which could communicate the information that the user needed, universally, understanding and manipulating design would require much less work. A way to communicate requirements of design processes as well as the way information transforms and flows throughout the design process is needed. Designers have product support design tools at their disposal. Using this as a starting point, a design tool based approach to modeling the design process could prove beneficial. Since design tools cannot compose a design process alone, the information which flows throughout the tools must also be considered. Using this approach, the information that enters and exits design tools could be connected to plot the design process evolution as it progresses.

Representing design processes with both product support and process evaluation tools could lend an advantage over other process models when representing design processes. With a model of design processes showing the importance and function of design tools, researchers could increase their understanding of design processes. Furthermore, the design process can be analyzed in individual portions as well as a whole to reveal fundamental relationships that are hidden within the design process.

A visual design tool based representation of design processes would enable researchers to modify the design process. When the design process is completely disassembled, the components and relationships within the process can be studied, thus allowing thorough understanding of the effect of each part of the process. This deep understanding could expose quantitative data about the effect of each design parameter and allow designers the ability to critique the design process. Having this concrete data would allow design process construction to be done on rules and facts which are easily quantified.

Observing and measuring precise rules and facts about design processes has been subjective for decades, but with the ability to "see" design, one could predict performance of designers engaged in design processes (Tufte, 1986). Engineers and designers are typically visually oriented people (Henderson, 1999). It is logical that if design processes could be illustrated with visual images, then the designers and engineers who use them so frequently would be able to better understand and manipulate the process to suit their needs.

Such a visualization of design should be capable of representing many if not any design process model. Researchers should be able to construct such a representation in context with their own domain of information. The users should be able to locate and follow any specified entity through the entire design process with ease. This model would enable performing visual and content based comparisons of multiple design processes. With such enhancements in design analysis, future researchers can build experiments to control design process outcomes. Such experiments would prove useful for design education as well as research projects. All of this could be applied to case study research, giving engineers a more powerful tool to advance the understanding of design. By observing the information flowing through each activity within the design process, bottlenecks can be spotted and addressed to streamline efficiency.

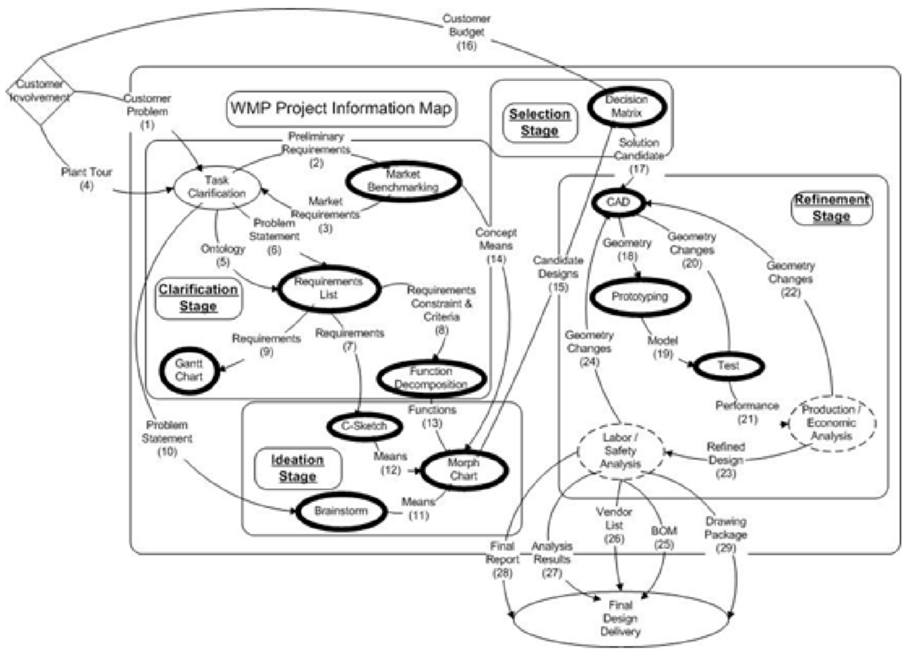
Fully understanding the design process and what happens within the context of design would allow engineers to improve the design process to suit the resources that are available to the designer. The design process itself could be manipulated intentionally to control the design process and learn about the effects that each component has on the final solution, thus blurring the line between experiments and case studies. With this improved understanding, the process can be used to improve the use of case studies in analyzing design processes. By implementing the aforementioned visualization method, case studies can become commonly used by many researchers and something that can work effectively to guide the design processes for engineers.

Thus, the development of a new process visualization tool is warranted. Here, we propose a visual representation scheme to illustrate information within design processes. This visualization method would enable the designer to see information that is of interest throughout the design process. As the process representation is constructed, the information of interest will develop from the first step of design to the final delivery of the product. Once complete, the process maps will allow researchers to follow critical information through the process as well as plan ahead to ensure that the current design process will yield something of sufficient value to the customer.

Development of a Design Process Visualization Tool for Case Study Research

A survey of process models is used to compare the differences between current design process model approaches and the limitations of each with respect to value gained by the executor. Inspired by IDEF0, an altered process model is presented in an attempt to increase the information captured by the designer when using and constructing these process models. The graphic representation proposed is applied to three case studies to illustrate the benefit of case study use in design tool development. Through applying this suggested representation to the case studies, it is shown that the Design Enabler Information Map (DEIM) allows researchers to create a visualization of design processes that can illustrate the needed information and be easily readable. By doing this, it is the goal of the representation to enable designers to understand, execute, analyze, and modify design processes in more effective ways than were previously available. By enabling design, the means of understanding, representing, modifying, and completing the process of design is permitted, or in this case, enhanced. Thus, this

Figure 5. Generic design enabler information map (WMP)



tool will be called Design Enabler Information Map (DEIM). Since information is the critical element of design, and the representation shows how information flows into and out of specific entities, these representations are called information maps. A design tool information map for the welding cell for metal frame construction project is shown below in Figure 5. Design activity sequencing is captured in the labeled arcs, along with the type of information that is flowing between design tools and activities. The design activities are grouped together into general design phases, such as *clarification stage*. The design activities are represented as formal tools (bold), general activity (thin), and unstructured (dashed).

Visualizing the design process traditionally falls into one of a handful of formats. Each of these formats has benefits and limitations to its use. Understanding what each can and cannot do is critical to proper application within the design process model. Fundamental methods to presenting the information contained within design process models exist and are commonly used widely; IDEF0, PERT, and Network schemes are typically the most commonly used (Dorador & Young, 2000; Shenoy, 2000).

Representation schemes for design processes are methods which present the information contained in the model to the reader. However, not all design process models are represented similarly. Some are represented via text and others are graphic with some being capable of using both. Words can be formed and sentences constructed to that communicate the information needed (Nanard & Nanard, 1995). While the text representation can give accurate and descriptive detail, the text is often lengthy and laborious to edit. Ideally, the detail will be maximized when writing the description and then filtered after the reader has deemed a specific portion unnecessary (Lockledge & Salustri, 2001).

A graphic representation may or may not give the same explicit detail as a text document, but can allow the user to observe the information that is pertinent to their present work. This can greatly reduce both the size of the representation as well as the volume of information that must be processed in order to obtain the required description of the image. A variety of graphic representations exist, and each performs a specific function appropriate for its own use. Graphics such as diagrams, pictures, and icons are the most abstract. These have the least structure and can appear in almost any different form. Other forms of graphic representations are charts, graphs, and plots. These often, but not always, relate directly to mathematical entities allowing great detail to be communicated in a relatively small form. Charts are typically used for comparison of few characteristics. Matrix or table representations also utilize visual stimulus to communicate. However, these combine the visual principal with traditional text to relay information. Understanding a matrix requires the reader to associate the row and column of the specified entity to the characteristics of that entity.

A comparison of Design Structure Matrices (DSM) to an image diagram in representing design processes has been evaluated against design research criteria (Keller, Eckert, & Clarkson, 2006). In this evaluation, the DSM is compared to a Node Link diagram. Both of these methods are used to predict changes in process modeling and are a form of connectivity modeling which in turn introduces coupling complexities (Summers & Ameri, 2008). In order to generate a method of representing design processes, a form of graphic representation must be selected; being either a matrix or node-link graph.

In evaluating matrices against node-link diagrams, measures needed to be established to give a level of goodness. The ease of locating a specific information unit or node is of interest. Likewise, identifying and locating a specific relationship, or link between two information units is also needed. The ability to count the number of links entering and exiting a node should be possible. The identification of any adjacent nodes should be available to the user with minimal effort.

It was discovered that matrices lend favor to sorting information, thus aiding in selection (Keller, Eckert, & Clarkson, 2006). Matrices are organized tables which can present large quantities of data in a clear fashion. However, matrices do make selection and counting of connecting links difficult. Identification of a relationship between two items is tedious to recognize meaning that identifying nodes adjacent to the current one is troublesome. Identifying the shortest process path via matrices is laborious if being analyzed manually.

Node-link diagrams present different benefits and challenges. They allow for identification of specific links easily. Determining the direction of the process links is easier with node-link diagrams, as is locating adjacent nodes. Following process paths is simple with node-link representations thus making the identification of the shortest possible path easy (Keller, Eckert, & Clarkson, 2006). The act of following paths, and shortest path identification were proven to be better represented with node-link diagrams for all datasets. This is believed to be due to the physical display of a path, which makes following easier. Since path following is critical to understanding information generation, flow, and change, node-link diagrams will be used to represent design processes.

IDEF0 is used to model functions of a system (Dorador & Young, 2000) and is similar to node-link diagrams. It represents the relationships and data which support the connection of those functions. The models are composed of hierarchal diagrams that show increasing levels of detail as the hierarchal level is reduced. IDEF0 is traditionally used to relate function to information of systems. The functions are traditionally shown as the nodes or entities. The information units or data which combine functions are shown as arrows leading from one function to another.

While IDEF0 diagrams allow designers to improve processes by describing critical information relationships, the traditional form of the tool is not optimal for representing process flows due to difficulty in tracking link flows. IDEF0 diagrams are suitable for showing activities and some connecting information. However, showing multiple domains of information flowing through nodes is something IDEF0 cannot easily communicate to the diagram reader.

An ideal scheme would combine the beneficial qualities of each of these into a single representation that researcher can use to illustrate design processes. The researcher should be able to specify the fundamental information of interest which is related to the specific information that flows through the process. The designer should also be capable of specifying the modifying bodies, or nodes, that exist so that the representation will be applicable to their specific area of work. The goal is to be able to determine the appropriate steps and expectations of design processes given the available information. The use of text should be available but not required and should not be limited to specific formats. This representation would be flexible to the researcher to use how each particular process needs it to in order to communicate the information that the designer desires to see.

The Case Studies

Three independent design projects were modeled after completion using the DEIM approach. The first project, designing a robotic welding cell, was a four month industry sponsored undergraduate design project for a small manufacturing firm. The second project was sponsored by another small company that is developing a combined trash and recycling collection and processing vehicle. The final project was to develop a tire testing system and was sponsored by a larger company. A brief comparison of the projects is found in Table 1.

These case studies are used to demonstrate the feasibility of using the DEIM to capture the type of information flow in the design projects. This quick and qualitative design research tool can allow engineers to visualize how different information is related throughout the design process. Coupling this visualization tool with formal ontologies for describing design information and activities is the next step in creating a formalism for conducting engineering design case study research.

For each of these projects, a visual design process map will be constructed representing the actual information and actions taken within the real life design project to develop a process representation and analysis tool. It is believed that throughout the development of these maps, the ability to represent design processes can be enhanced. Through these projects, the maps will be developed and tested, to show that visualization of design processes is possible and beneficial to researchers.

The graphic representation proposed is applied to three case studies to illustrate the benefit of case study use in design tool development. Through applying this suggested representation to the case studies, it is shown that DEIM allows researchers

Table 1. Case Study Comparisons

	EAI	WMP	Michelin
Duration	24 Months	4 Months	14 Months
# of Participants	6	6	8
Review Frequency	4 Weeks	1 Week	1 Week
Professors	1	0	2
Graduates	4	0	5
Undergraduates	1	6	1
Deliverable	Revised Vehicle CAD	New Work Cell	New Equipment and Testing Procedure
Time from project to mapping	5 Months	20 Months	1 Month

to create a visualization of design processes that can illustrate the needed information and be easily readable. It is the goal of the representation to enable designers to understand, execute, analyze, and modify design processes in more effective ways than were previously available. By enabling design, the means of understanding, representing, modifying, and completing the process of design is permitted, or in this case, enhanced. Thus, this tool will be called Design Enabler Information Map or DEIM. Since information is the critical element of design, and this representation shows how information flows into and out of specific entities, these representations are called information maps.

By intent, DEIM is to promote illustrative flexibility in what is represented as well as how it is represented for the case studies modeled. The main consideration when building a DEIM is to structure each connection in the context of what is being tracked, the information of interest. However, some formalism is required to enable logical perception by readers (Kim & Jang, 1999). The ability to freely position nodes and links within the map remains, but the net flow of the process should extend from the top and left to the bottom and right of the map. Additionally, the method of representing information types cannot be shared by multiple types. These are requirements that permit the process to be modeled flexibly yet with organization.

DEIM's consist of a few components which are connected together to create the representation of the process. These are the information links, the accompanying nodes, and any specific borders which are an extension of the nodes. Information is represented in the form of arrow links, which may have specified thickness, color, line style, length, labels, and direction (Kim & Jang, 1999). The nodes can consist of various shapes, sizes, colors, and labels. Any graphical property of the maps can be defined as a characteristic key of the map, thus representing some specific information unit. Borders shown within DEIM perform the same function of nodes, establishing detail limits to the investigation of the process. Borders are shown on the highest level as the design process boundary allowing initial information to enter and the solution information to exit.

Each component in DEIM represents some sort of information that is important to the overall representation. Information links represent the most elemental component of the process representation. The nodes represent the transformation of the information from one state to another. The borders that exist in the DEIM are expansions of higher level nodes. They serve to group information links and action nodes of specific relation together.

The goal of constructing a DEIM is to sufficiently track the initial information given to the designer through each step it is transformed through until the desired exiting information is reached. This will allow the designer to understand how information changes, how information flows, and how information is related to specific operations within a process.

DEIM can be used to represent actual complex design processes, consisting of a multitude of information domains, and including an assortment of active entities which work together to illustrate a design process in a clear and concise manner. The information contained within a DEIM can be low if the mapper desires, resulting in a DEIM showing a single type of information throughout the process. Additionally, DEIM can show many different types of information while still remaining readable by the user. Ultimately, they show what happens in a process as well as how it happens by illustrating the connectivity and transformation of links through nodes.

To test the use of and verify the benefit from DEIM, three design projects were represented using the DEIM method in a manner similar what others have used in validating their case study work (Hernandez, et al., 2001). For the DEIM case studies, the author was involved with each of these projects thus lending firsthand experience to the events as they occurred, giving the author the ability to map the events of the design processes into DEIM. These DEIM models were constructed after the project was complete, although the time span between project completion and map development varies for each example shown.

For each of the examples shown, the information links are generic design process information that may range from Requirements Lists to Concept Performance. A formal ontology for the type of information exchanged between the design tools has not yet been adopted. The nodes selected are the combination of design methods and tools (instantiations of tools). The author recognizes the difference in these two, but the distinctions are not of interest, therefore they will be used collectively. These serve to transform the information from its entering state to some other form via work.

In these maps, several generic design stages exist such as, Clarification, Ideation, Selection, and Refinement. These were chosen based on a traditional systematic engineering design process (Pahl & Beitz, 1996). Each respective map may or may not show the explicit stage boundaries, depending on the type of work being done. An example of this is the EAI project which consisted of three separate and concurrent detail refinement stages. The baler, trash compactor, and structure systems all were designed independently, at the same time. Rather than show a single refinement stage, the EAI map shows each separate system refinement stage of the design process.

THE EAI PROJECT

The Environmental America Inc. (EAI) project was a privately sponsored endeavor. The company, EAI, had an idea to patent and prototype an integrated recycling center on board a common trash collection truck. EAI approached designers with a fourth generation prototype and asked the researchers to "streamline" the truck by reducing cost, mass, and complexity. This project relied heavily on the task clarification stage of design. Significant work was done to justify implementation of such a product into the infrastructure of local cities and residual plans had to be made that would affect the performance of the end product. Once appropriate justification of such a device was found, work began on ideation of concepts and product breakdown. The fundamental requirements of the truck were developed and used to eliminate excess systems on the current prototype. Once certain systems were eliminated, work began to improve efficiency and performance of the needed systems to conduct the truck's tasks. This work was segmented into three areas; Trash collection, Baler Design, and Structure Design (Smith, Johnston, & Summers, 2007).

The first of these is the design of the Trash handling equipment. This involved conceptualizing equipment to compact the trash and eject the load once at the land fill. Significant difficulty was encountered in modeling the behavior of trash, therefore, the project funding ceased before sufficient information about this could be collected. The baler design was more productive. The testing phase of this design proved instrumental in the generation of both concept ideas and Loading Conditions for the simulation. The superstructure design was also completed. Specification information was gathered from benchmarking existing vehicles and reproducing the functionality of those designs. The safety of the concepts was checked via FEA and the loadings gathered from market research.

Representation of the EAI Project

The DEIM of the EAI project was completed five months after the design project was finished. The DEIM representation of the EAI project is shown below in Figure 6. The information links vary in content but are primarily generic process information. Notice that each of the nodes is a design tool or design method which was used to transform the information to some other form. The line style of the node signifies what that node is. A bold node outline represents formal design tools. A thin node outline represents an informal action which can possibly be a design tool or method. Finally, the dotted outlines represent a fuzzy definition of that particular node. For these, the intent may be common, but the way in which the action was taken may constitute some modified version of a standard task definition. Thus these types of nodes are classified as "Fuzzy" and shown with a dotted line.

For this example, the information links can be related back to physical design documents or files. The Requirements List is a document that was printed and presented to the client. The geometry generated from the solid modeling (SolidWorks) step is the *.prt file for each respective part that was then analyzed in finite element analysis (Cosmos).

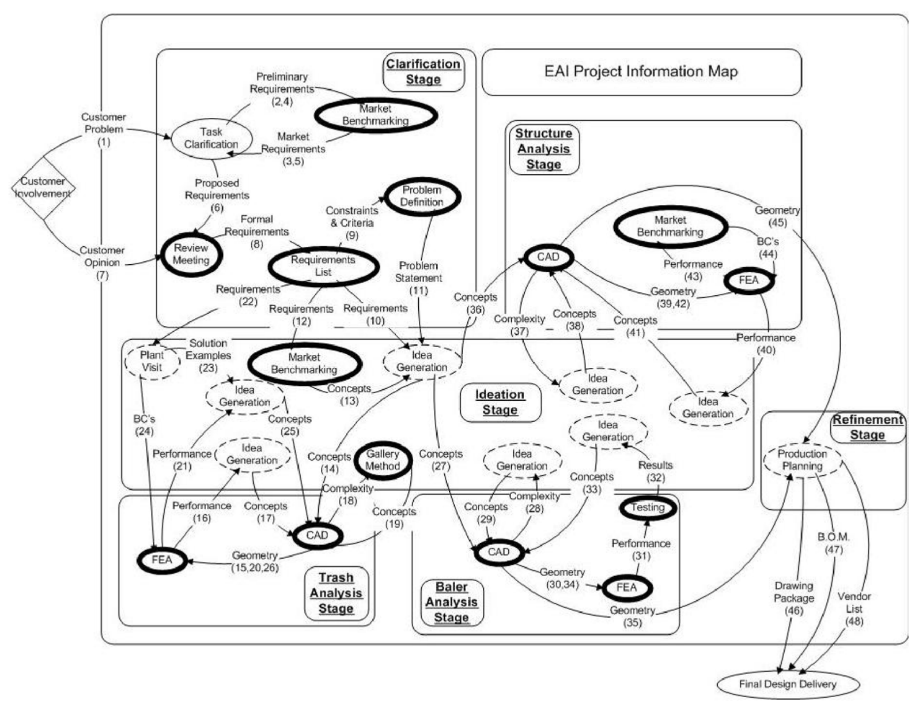
Observations of EAI DEIM

By constructing and examining the EAI DEIM, it can be concluded that:

- DEIM construction iteration improves readability
- Existing labeling of DEIM does not support parallel node sequencing
- DEIM readability can benefit from using cognitive tendencies
- DEIM construction iteration improves process content accuracy
- DEIM enables researchers to identify non-value adding nodes within a process
- DEIM enables researchers to identify information critical to a specific process

DEIM can allow the researcher to identify the critically needed steps within a process. By starting at the final design delivery, each subsequent step can highlighted. This can propagate all the way

Figure 6. Final EAI DEIM



up to the initial customer input. By doing this, connection of each part of the DEIM to the final design delivery can be determined. In doing so, non-value adding steps are identified, revealing to researchers the wasteful steps within the process.

By studying the following maps, Figure 7, Figure 8, Figure 9, and Figure 10, it can be seen that the Trash Analysis Stage along with Idea Generation Steps connected to it produced no work that contributed to the final design. This occurred because the effort required to model compacted trash behavior is extensive and the client decided to purchase an existing unit rather than consume resources to develop a new one. The identification of non-value adding nodes begins in Figure 7, with the final product. The required input information to that node is identified and then the source nodes of that information are then highlighted. This process propagates back through the process until the initial step is reached. Once a complete path of value adding nodes reaches from the final solution to the initial link, the process evaluation is complete. At this point, any remaining nodes are non-value adding.

THE WMP PROJECT

The Wright Metal Products (WMP) project was a four month long, sponsored design project. The client had purchased a welding robot, and wanted the designers to develop an efficient work area to use the robot in their facility. The client manufactures steel shipping crates out of pre-cut parts which are then welded manually in the facility. The goal of this project was to improve efficiency and production by using the robot along with manual workers. The client gave freedom to the designers to specify all needed layout, equipment, and fixtures that would be needed.

Figure 7. Back propagation step 1

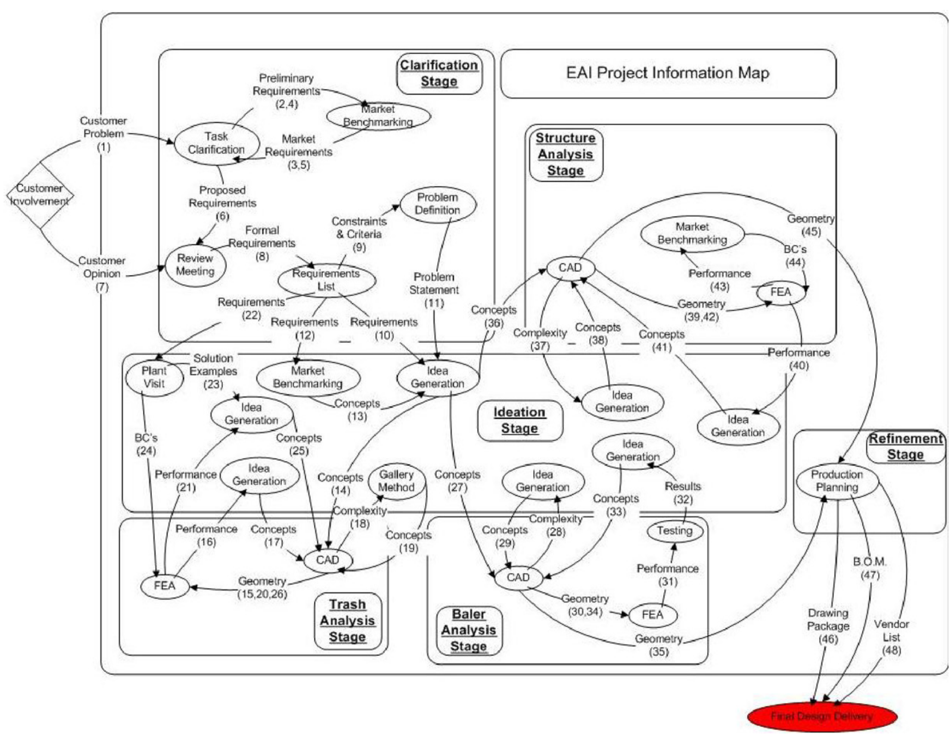


Figure 8. Back propagation step 2

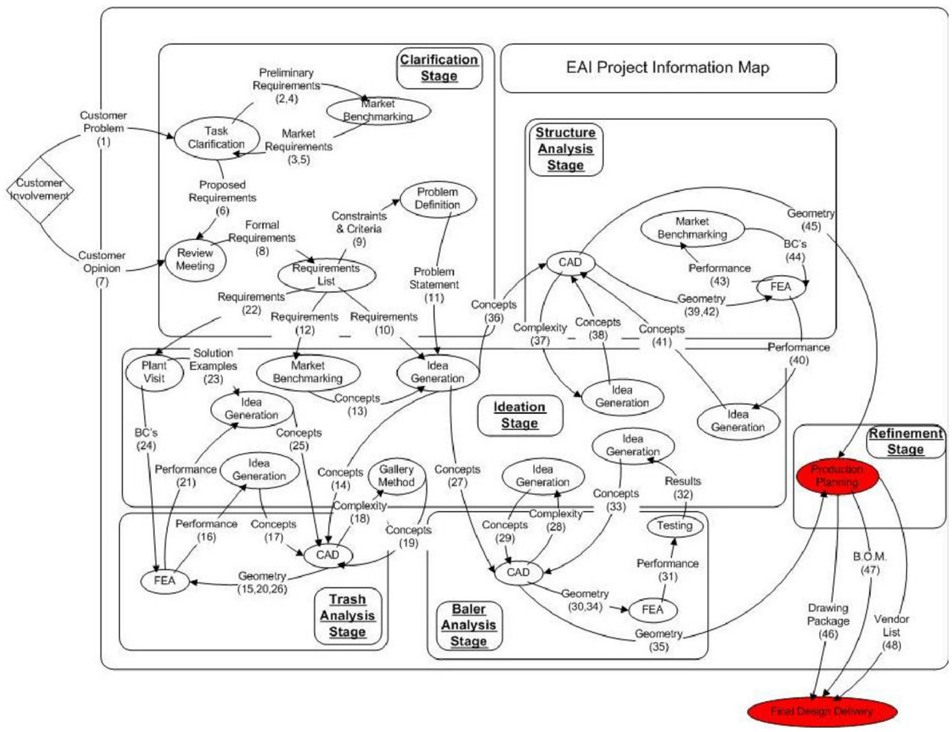


Figure 9. Back propagation step 3

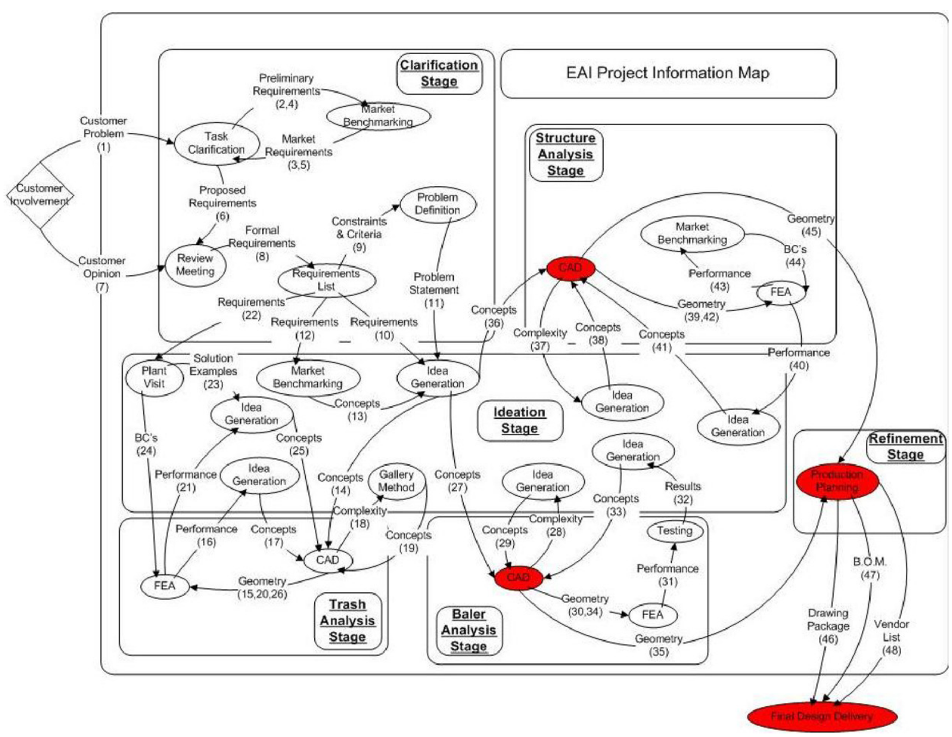


Figure 10. Back propagation complete

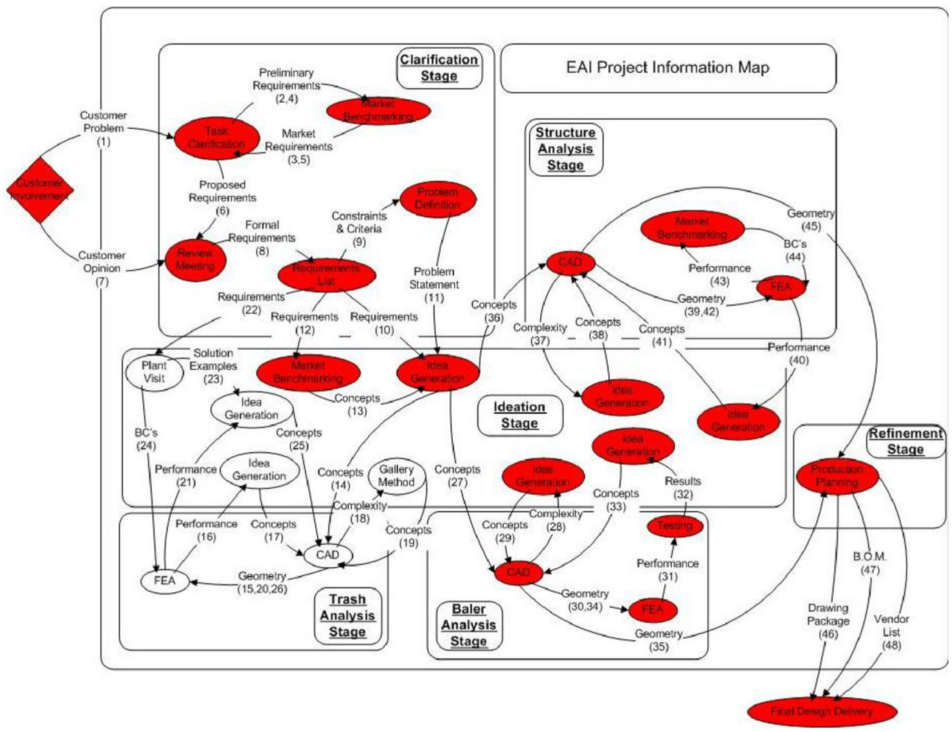
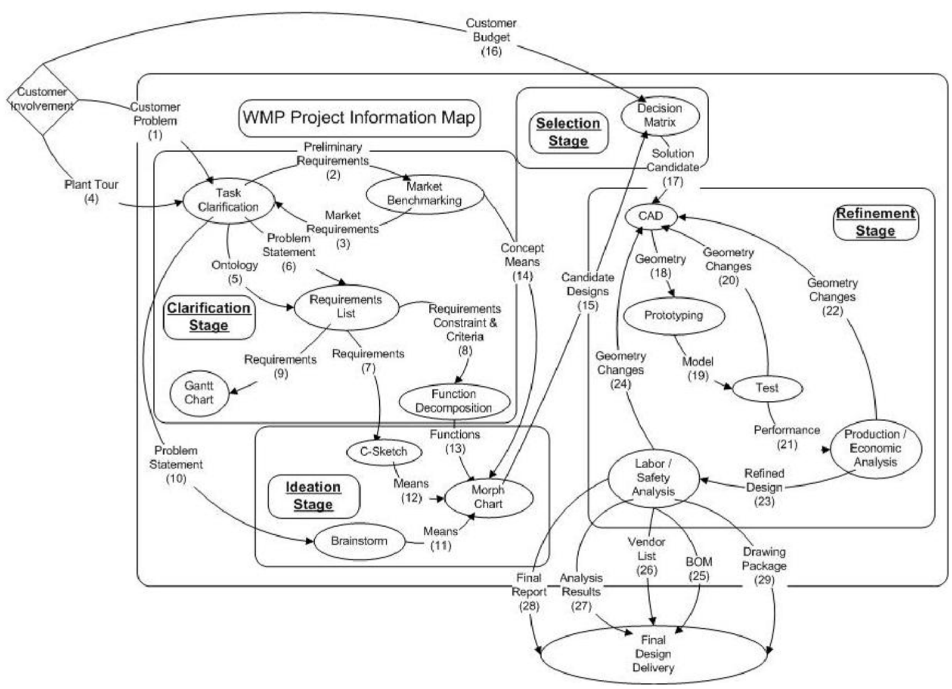


Figure 11. WMP DEIM



Representation of the WMP DEIM

The DEIM for the WMP project is shown below in Figure 11.

It can be seen that this map does not contain the bold, thin, and dotted node outlines. That representation scheme was used in the EAI map to differentiate the different classifications of nodes that were used. For the WMP map, most of the nodes were formal design tools, therefore adding the different node formats would only serve to add information to the map which is not desired.

This project consisted of four main stages, starting with the Clarification, proceeding to the Ideation, then to the Selection, and completing the process with the Analysis. The Clarification stage served to properly define the problem and develop forma Requirements that must be met by the solution. The Ideation stage encompassed three idea generation techniques and the formation of a Morphological Chart to aid in concept organization. The Selection stage was the shortest in duration and consisted of the use of a Decision Matrix to rank solution candidates allowing the designers to choose the better concepts from all the ones generated. The final stage was the Analysis stage, where the solution was tested, analyzed, refined and validated into a suitable solution.

Observations

By constructing and examining the WMP DEIM, it can be concluded that:

- DEIM is flexible with respect to node classification and formality
- DEIM is flexible with respect to node grouping
- DEIM can enable researchers to identify non-value adding nodes
- Nodes can alter information and channel information

Some of the nodes in the WMP map accomplish dual functions. The first is that the node functions as the modifier of the information. Initially, this feeds back into the previous steps for iteration. However, when the information again reaches the same node, a different "check" may occur, thus allowing the information to "pass" through the node on to the next. An example of this occurrence is shown below in Figure 12.

THE MICHELIN PROJECT

The Michelin project was a twelve month long project that began when representatives from Michelin Tire Co. approached the engineers with the topic of test procedure development. They were interested in understanding a soil and tire interaction and wanted to investigate the possibility of the designers developing an economic solution to their need of testing.

Representation of the Michelin DEIM

The DEIM of the Michelin Project is shown below in Figure 13.

Tires are a complex machine, and the construction and simulation of their behavior is not a

Figure 12. Dual function nodes

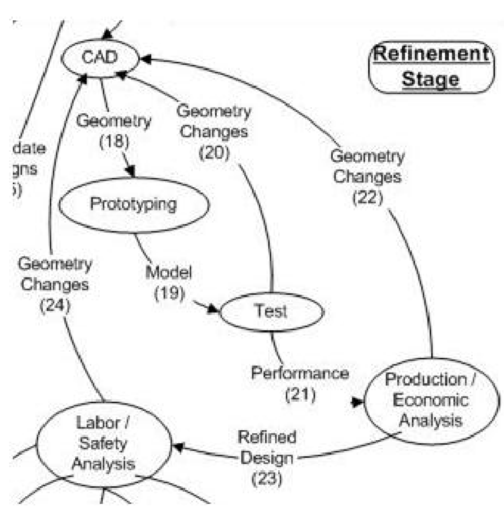
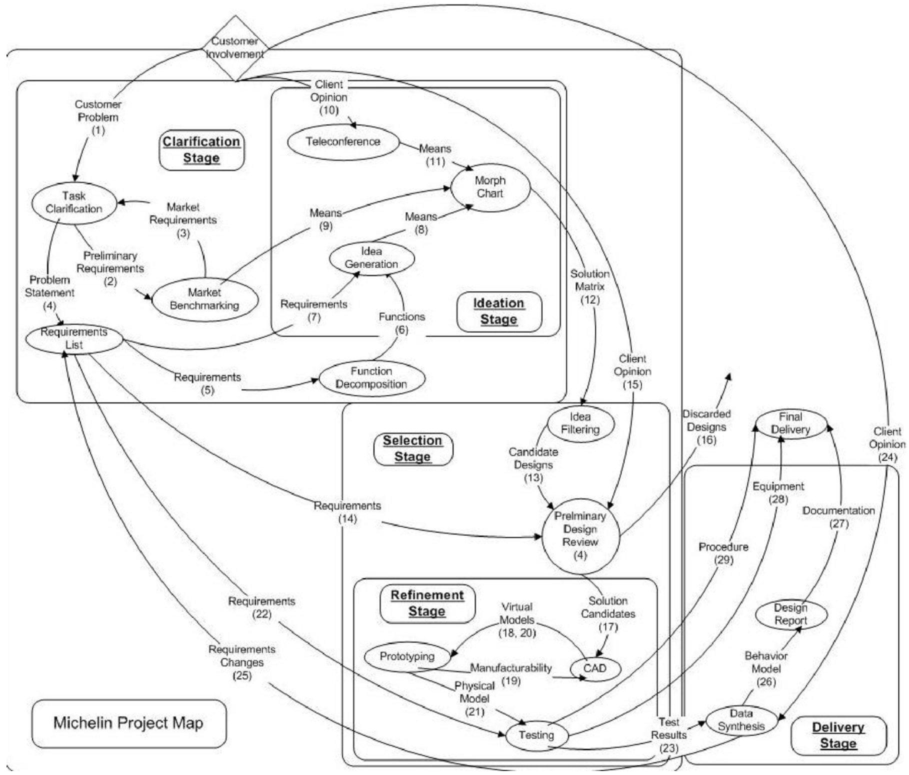


Figure 13. Michelin DEIM



simple task. To test specific behaviors, tires are generally constructed, and used to examine the performance of certain parameters. The engineers will then specify changes, and will need to manufacture a new set of tires. This is costly due to tire mold manufacturing cost. The testing cost limits the ability of the tire engineers to learn about the behaviors, so the need of a simulation protocol with more economic operational cost is warranted. This was the goal of the designers, as specified by the customer.

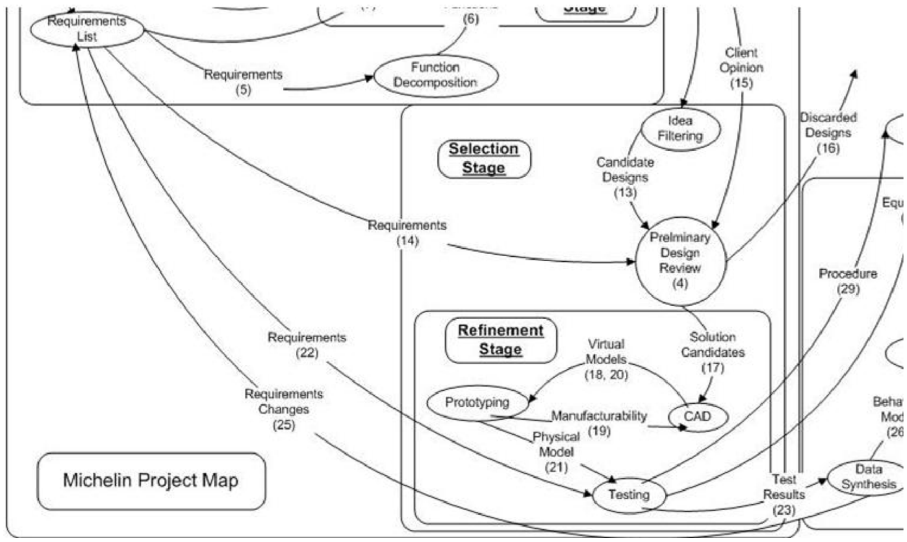
Observations

By constructing and examining the Michelin DEIM, it can be concluded that:

- DEIM can show non-critical information if the mapper wishes
- DEIM is flexible with respect to node grouping
- Link length may be proportional to specific information properties
- Node connectivity may be proportional to specific information properties
- DEIM can show the result of insufficient information
- DEIM can show the effect of making long iterations

It can be seen with this map (Figure 14) that the length of the links is proportional to the goodness of the design process. As mentioned above, early and often iterations make for better design processes, whereas the links representing those iterations would be spaced close together and in stages rather than in long sweeping curves. The link length in this DEIM varies greatly from node to node. This communicates some information

Figure 14. Michelin DEIM link length



about the process such as time required to iterate and percentage of time wasted by the iteration.

The Michelin DEIM shows relationships of node connectivity to the difficulty of that part of the process. If a node is highly connected such as the Requirements List or the Customer Involvement, then accurate and precise information from that source is needed for the design process to proceed properly. Accordingly, incomplete or inaccurate information from highly connected sources works against the productivity of the design process. As mentioned above, the lack of exhaustive requirements by the customer hindered the progress of the design project. If the significance of the Requirements List been known initially, the designers may have urged further clarification and development of the Requirements List before proceeding on to the next step of the design process.

The Michelin DEIM also shows the effect of insufficient information being used in a process. Fault of this deficiency cannot be given, but the effect of completing most of the process without it can be shown with the map. As the design data is being synthesized, discoveries were made that amended the requirements of the process. DEIM shows this by connecting a late node to an early node, thus initiating an almost complete process repetition.

CONCLUSION

By representing information within design processes, researchers can model complex processes graphically. Doing so gives researchers a better mental connection between what is perceived and what is real. With better understanding comes improved ability to analyze design processes with case studies. These improvements were shown through the application of a newly developed design process model in three case studies. Each case consisted of different deliverables, different experience levels of the designers, and project durations. The use of case studies to develop design tools and conduct design research has been proved beneficial by the successful application of and evaluation with DEIM to the case studies.

Through conducting these case studies, observations about the DEIM representation of design processes can be made and are:

- DEIM construction iteration improves readability
- Existing labeling of DEIM does not support parallel node sequencing
- DEIM readability can benefit from using cognitive tendencies
- DEIM is flexible with respect to node classification and formality
- DEIM is flexible with respect to node grouping
- DEIM can show non-critical information if mapper wishes
- DEIM is flexible with respect to node grouping

These observations illustrate ways that researchers can benefit from a graphic representation of design processes and that design processes can accurately be represented graphically while containing sufficient detail and promoting readability.

Observations about the cases studied through the application of DEIM were also made here are:

- DEIM construction iteration improves process content accuracy
- DEIM enables researchers to identify information critical to a specific process
- DEIM can enable researchers to identify non-value adding nodes
- Nodes can alter information and channel information
- Link length may be proportional to specific information properties
- Node connectivity may be proportional to specific information properties
- DEIM can show the result of insufficient information
- DEIM can show the effect of making long iterations

Information paths can be easily traced with DEIM. From the initial problem of the client to the documentation or artifacts that are physically delivered to the client, the type, quantity, and change of the information present at any given stage of the process can be displayed. This gives designers an ability to determine "Next Steps" for each process as well as determining information that may be needed to complete a specific task.

The critical path of information flow can be determined from reading DEIM. This can be done via back tracking the preceding nodes or any other means desired. The representation of links and nodes allows a holistic understanding of the process to advance into a detailed understanding of information that enters and exits any specific node.

The connectivity of the nodes within a DEIM can also be seen. It was shown that DEIM nodes that are highly connected tended to be significant contributors to the design process. The concept of this metric can be used to build DEIM in respect to other characteristics such as production dependencies, or information management. In the case studies shown, the connectivity relates to both the affect that the particular node has on other nodes as well as the difficulty that can be encountered when constructing DEIM containing that specific node.

More importantly, the development, application, and evaluation of new design tools and methods through the use of case studies has been proven. The three cases used show the benefit of the developed representation through the information revealed about the cases. The cases served to test the new representation, and allow the author to discover information about the representation that was previously unknown. These observations have been mentioned in the previous section. Furthermore, the application of the new tool to case studies has also shown that case studies can be used to evaluate design processes, in attempt to understand the process meaning that future endeavors of modifying and improving the process are possible.

Certain limitations were encountered when using DEIM to represent tools used in processes. These difficulties came from constructing the DEIM as well as reading and understanding the process being represented. The first difficulty is the difficulty to search for specified objects (links or nodes). With other representations such as DSM, readers can search for a specified entity in a one dimensional, linear manner by scanning either a column or row heading. DEIM expands the representation into a full two dimensional space, thus making searching more difficult. This was addressed by Keller et al as well in their Matrix vs. Node Link evaluation (Keller, Eckert, & Clarkson, 2006).

The layout of the nodes is also a cumbersome and tedious process. It is difficult to position the nodes of highly connected processes in a manner such that the links do not intersect. In fact, sacrifices must sometimes be made when constructing the maps to show some information and alter others. DIEM is a user defined representation of the process. It could be beneficial to allow link intersection sometimes, thus creating another entity which can represent additional information about the process.

The work presented in this chapter leads to some areas of future investigation such as the exploration of domain representation methods. They ways in which information domains can be shown within a map cannot currently be counted. The ability of DEIM to illustrate the sequence of operations within a design process is also of value to researchers. Future work could be focused on how to represent time and process sequence within the DEIM. DEIM contains complexity that could greatly benefit from computer integration. Implementing the theory presented in this chatper into a software package could prove to generate a useful and powerful tool for designers. Design disciplines such as project management could use such a program in researching and developing their own principles. More case studies should also be completed using DEIM to model the processes. These case studies give specific context to issues revealed in each map and allow researchers to relate the fundamental elements of DEIM to the details of design processes. The theory of information mapping reveals issues that could prove beneficial to process representation schemes. The application of the Conservation of Energy principle into information mapping could prove beneficial. By theorizing that "information cannot be created nor destroyed, but can only change form", researchers may be able to enhance the way that design processes are understood and completed. Similar benefit may also be had by investigating the integration of Boolean Logic into information maps. By identifying non-value adding steps, DEIM opens opportunity of integration with lean manufacturing process models. This can also be an extension of a software package. By exploring the theory of both lean manufacturing and DEIM, researcher may be able to discover similarities, thus enabling the design community to gain a broader and applicable model to understanding processes.

Finally, the expansion of design process models away from verbal descriptions into visually grounded renderings should be continued. This research has formed two dimensional representations of complex and real world design processes. A lofty future goal of this research is to enable designers to use complex and detailed scientific visualization tools similar to CAD software packages to represent design processes in a multi-dimensional virtual environment. This would give greater capability and freedom in information mapping and process modification to the researcher by representing design processes in virtual three dimensional spaces.

REFERENCES

Ahmed, S. (2007). An Industrial Case Study: Identification of Competencies of Design Engineers. *Journal of Mechanical Design*, 709–716. doi:10.1115/1.2723807

Atkinson, R. (1999). Project Management: Cost, Time, and Quality; Two Best Guesses and a Phenomenon, Its Time to Accept Other Success Criteria. *International Journal of Project Management*, 337–342. doi:10.1016/S0263-7863(98)00069-6

Battista, G. D., Pietrosanti, E., Tamassia, R., & Tollis, I. G. (1989). *Automatic Layout of PERT diagrams with X-PERT* (pp. 171–176). Washington, DC: IEEE.

Burge, J., & Brown, D. (2002). Integrating Design Rationale with a Process Model. *Workshop on Process Modelling, AID'02: International AI in Design Conference.* Cambridge, UK.

Dorador, J. M., & Young, R. I. (2000). Application of IDEF0, IDEF3, and UML methodologies in the creation of information models. *International Journal of Computer Integrated Manufacturing*, 430–445.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532–550. doi:10.2307/258557

Gero, J. S. (1990). Design Prototypes: A Knowledge Representation Schema for Design. *AI Magazine*, *11*(4), 26–36.

Grabowski, H., Lossack, R. S., & Weis, C. (1995). A Design Process Model based on Design Working Spaces In *T. T., M. M., & S. Finger, Knowledge Intensive CAD-1* (pp. 26–29). New York: Springer.

Hazelrigg, G. A. (1998). A framework for decision based engineering design. *Journal of Mechanical Design*, 653–658. doi:10.1115/1.2829328

Henderson, K. (1999). *On Line and on Paper: Visual Representations, Visual Culture, and Computer Graphics in Design Engineering*. Boston: MIT Press.

Hernandez, G., Allen, J. K., Woodruff, G. W., Simpson, T. W., Bascaran, E., & Avila, L. F. (2001). Robust design of families of products with production modeling and evaluation. *Journal of Mechanical Design*, 183–190. doi:10.1115/1.1359786

Keller, R., Eckert, C. M., & Clarkson, P. J. (2006). Matrices of node-link diagrams: which visual representation is better for visualizing connectivity models? *Information Visualization*, 62-76.

Kim, S.-H., & Jang, K.-J. (1999). Designing Performance Analysis and IDEF0 for Enterprise Modelling in BPR. *International Journal of Production Economics*, 121–133.

Kirschman, C. F., Fadel, G. M., & Jara-Almonte, C. C. (1996). Classifying Functions for Mechanical Design. *Design Engineering Technical Conference.* Las Vegas, NV: ASME.

Kumar, P., & Mocko, G. M. (2007). Modeling and Analysis Of Engineering Design Activities Ontology Using the Design Structure Matrix. *Design Engineering Technical Conference.* Las Vegas, NV: ASME.

Leifer, L. J., & Tang, J. C. (1988). *A Framework for Understanding the Workspace Activity of Design Teams*. New York: ACM.

Lockledge, J. C., & Salustri, F. A. (2001). Communication Using a Variation of the Design Structure Matrix. *International Conference on Engineering Design.* Glasgow: ICED.

Maylor, H. (2001). Beyond the Gantt Chart: project management moving on. *European Management Journal*, 92–100. doi:10.1016/S0263-2373(00)00074-8

Nanard, J., & Nanard, M. (1995). Hypertext design environments and the hypertext design process. *Communications of the ACM*, 49–56. doi:10.1145/208344.208347

Ostergaard, K., & Summers, J. (2007). Resistance Based Modeling of Collaborative Design. *Concurrent Engineering Research and Applications*, *15*(1), 21–32.

Pahl, G., & Beitz, W. (1996). *Engineering Design: A Systematic Approach*. New York: Springer-Verlag.

Reymen, I. M., Hammer, D. K., Kroes, P. A., Aken, J. E., Dorst, C. H., & Bax, M. F. (2006). A domain-independent descriptive design model and its application to structured reflection on design processes. *Research in Engineering Design*, *16*(4), 147–173. doi:10.1007/s00163-006-0011-9

Shenoy, P. P. (2000). Valuation network representation and solution of asymmetric decision problems. *Eurobean Journal of Operational Research*, 579-608.

Sim, S. K., & Duffy, A. H. (2003). Towards an ontology of generic engineering design activities. *Research in Engineering Design*, 200–223. doi:10.1007/s00163-003-0037-1

Simon, H. A., Kotovsky, K., & Cagan, J. (2001). *Formal Engineering Design Synthesis*. New York: Cambridge University Press.

Smith, E., Johnston, P., & Summers, J. D. (2007, September). Applying Lean Manufacturing Principles to Revolutionize Curbside Equipment and Collection Process. In E. Smith, P. Johnston, & J. Summers (Eds.), ASME-DETC2007, DFLM-35615. Las Vegas, NV: ASME.

Sobek, D. K., II. (2007). From Journals to Validated Tool: Results of Empirical Research on Student Designers. *International Conference on Engineering Design.* Paris: ICED.

Summers, J. D., & Ameri, F. (2008). *An algorithm for assessing design complexity through a connectivity view. Tools and Methods for Competitive Engineering*. Izmir, Turkey: TMCE.

Teegavarapu, S., & Summers, J. D. (2008). *Case Study method for design research. Design Engineering Technical Conferences*. New York: ASME.

Tufte, E. R. (1986). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Ullman, D. G., Dietterich, T. G., & Stauffer, L. A. (1988). A Model of the Mechanical Design Process Based on Empirical Data. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *2*(1), 33–52. doi:10.1017/S0890060400000536

Ulrich, K. T., & Eppinger, S. D. (2008). *Product Design and Development* (4th ed.). New York: McGraw-Hill.

Whitfield, R., Duffy, A., Coates, G., & Hills, W. (2002). Distributed Design Coordination. *Research in Engineering Design*, *13*(4), 243–252.

Yang, M. (2007). *Design Methods, Tools, and Outcome Measures: A Survey of Practioners. ASME International Design Engineering Technical Conferences (pp. DETC2007-35123)*. Las Vegas, NV: ASME.

Yin, R. K. (1994). *Case Study Research: Design and Methods* (2nd ed.). Newbury Park, CA: Sage Publications.

This work was previously published in Handbook of Research on Trends in Product Design and Development: Technological and Organizational Perspectives, edited by Arlindo Silva and Ricardo Simoes, pp. 80-105, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 87
Application of Dynamic Analysis in a Centralised Supply Chain

Mu Niu
Northumbria University, UK

Petia Sice
Northumbria University, UK

Ian French
University of Teesside, UK

Erik Mosekilde
The Technical University of Denmark, Denmark

ABSTRACT

This chapter presents dynamic analysis of a model of a centralised supply chain. The research was conducted within the manufacturing sector and involved the breathing equipment manufacturer Draeger Safety, UK A simplified model of the Draeger Safety, UK centralised supply chain is developed and validated. Simulation and analysis are performed using System Dynamics, control theory, nonlinear dynamic and chaos theory. The findings suggest that destructive oscillations of inventory could be generated by internal decision making practices. Bifurcation diagram is plotted to indicate the bifurcation status of the model with different internal decision policies. A management microworlds is developed for managers to experiment with different decision scenarios and learn how the supply chain performs.

INTRODUCTION

The contemporary business environment is characterised by an acceleration of the process of change due in part to improved communication and analytical capabilities brought about by modern Information and Communication Technologies (Sterman, 2000; Terzi and Cavalieri, 2004; Ortega and Lin, 2004). This poses an ever increasing demand for periodical review of strategies and organisational processes, and for faster and more effective learning to deal with problem situations more quickly. Furthermore, companies increasingly need to react to customer requirements (Hong-Minh, Disney and Naim, 2000). Naylor et al (1999) suggested the principles of

DOI: 10.4018/978-1-4666-1945-6.ch087

agile manufacturing, i.e. responding to customer requirements with very short lead-times can be applied to supply-chains. Global companies are realising that the efficiency of their businesses is dependent on the collaboration with their suppliers and customers (van der Zee, 2005; Yusuf, Gunasekaran, 2004; Adeleye and Sivayoganatahn, 2004; Cox et al, 2004). The concept *of supply chain management* (SCM) is concerned with the strategic approach of integrating producers, suppliers and customers in a collaborative network with the objective of improving the overall responsiveness of the network, but at the same time preserving the organisational autonomy of each unit (Terzi and Cavalieri, 2004). More accurate data and better information available to all nodes of the supply chain are considered crucial in improving its performance. Thus, companies invest heavily in intra-company information and communication platforms and E-collaboration tools, such as data warehouses or Enterprise Resource Planning (ERP) and Manufacturing Resource Planning systems (Lyneis, 2005; Li & Li, 2005; Marquez, Bianchi and Gupta, 2004).

Contemporary Manufacturing Resource Planning (MRP) is a structured approach to manufacturing management in which a suite of integrated computer software covering the main operational business functions is used to assist the planning of materials, capacity and cash flow, in accordance with company policy, to meet customer delivery requirements (Ackermans & Dellaert, 2005). It is a total approach to managing a business (Li & Li, 2005). It includes inventory control data, purchasing and accounts payable, suppliers' data, master schedules, production schedules, resource capacities, manufacturing orders, and so on. MRP systems are developed under the assumption that more data and better information lead to better decisions. The quality of this data is believed to be accurate or at least good enough to be reliable (Ackermans and Dellaert, 2005). However, even with the introduction of resource planning systems, the performance of the supply chain often remains problematic (Lyneis, 2005; Fowler, 1999). The weakness of such systems is that when used on their own they fail to address the complexity of management situations for two reasons: firstly, they do not take into account the inherent 'messiness' of situations that contain human decision making within the process that is considered problematic (Maturana & Varela, 1987; Beer, 1979, 1994); secondly, such tools do not promote learning (Senge, 1984, 1996; Venix, 1994; Morecroft, 2007) or effective decision support as they do not include simulation facilities to allow what-if analysis on alternative strategies (Terzi and Cavalieri, 2004).

After reviewing more than eighty scientific papers on the role and application of simulation in the supply chain, Terzi and Cavalieri (2004) show how simulation is successfully adopted in different studies of logistic networks. Their findings ascertain that simulation allows evaluation of performance prior to the implementation of the system since: 1) it enables companies to perform powerful what-if analyses leading to learning and improved planning decisions; 2) permits the comparison of alternative strategies in a safe virtual world, without interrupting the real system; 3) permits time compression that allows for timely policy decisions to be made. These findings are also supported by Chang and Makatsoris (2001).

The suitability of simulation as an approach to studying supply chains does not guarantee its adequate decision support (van der Zee, 2005) as simulation relies on a heuristic search for good decisions strategies led by people. The success of the simulation study depends on the joint work of the analyst and the chain managers and actors, as well as the facilities offered by the modelling and simulation tools. Van der Zee (2005), Sterman (2000) and Senge (1996) suggest that it is important to consider the simulation model as a communicative tool between analysts and chain actors. System Dynamics (Forrester, 1961; Sterman 2000) as a modelling and simulation framework has been developed for such a purpose.

This research uses a combination of System Dynamics, control theory and bifurcation analyses to address the issues of decision support and learning in the supply chain. It acknowledges the insights of non-linear dynamics and chaos theory in exploring the system's behaviour (Sice & French, 1998; Mosekilde, 2002).

When used in a management context *Systems Dynamics* is 'a framework for thinking about how the operating policies of a company and its customers, competitors and suppliers interact to shape the companies performance over time' (Forrester, 1961). It is an approach that allows practitioners to explicitly model perceived delays, bounded rationality and goal setting and explicitly discuss systems' boundaries (Ackermans and Dellaert, 2005). To do this System Dynamics uses metaphors drawn from control theory to provide tools for mapping and simulation, and metaphors from behavioural decision theory to describe decision making processes (Morecroft, 1985, 2007).

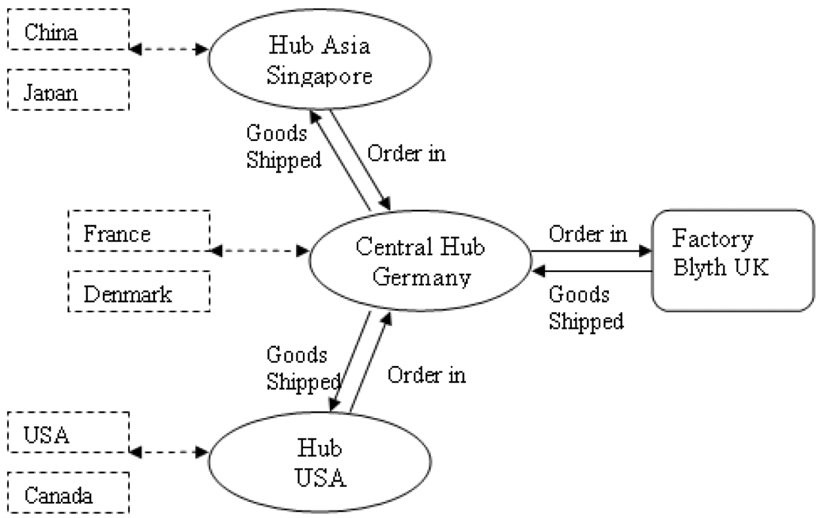
Systems Dynamics is a powerful framework for thinking and explaining how a system's structure generates behaviour. Ortega and Lin (2004), after reviewing the critical literature on System Dynamics conclude that this approach is suitable for modelling complex systems but when it comes to analysis, it contains insufficient analytical support. Therefore, if we would like to carry out in-depth mathematical analysis of behaviour we need to introduce appropriate mathematical tools from control theory (Ackermans and Dellaert, 2005). System Dynamics was founded in control theory and thus, is compatible with its tools and techniques. Ortega and Lin (2005) after studying the use of control theory in supply-chain analyses suggest that control theory analytical tools are quite suitable. However, in the supply chain there is an asymmetric consequences, for example, negative and positive inventories are asymmetric and thus introduce non-linearity in the system. Recent research suggests that non-linearity in the supply chains may lead to chaotic behaviour (Thomsen et al, 1992; Mosekilde, 2002; Sice, French et al, 2000). The control theory is generally linear in nature, and in order to consider effective application in supply chain analysis, it needs to be complemented by non-linear dynamics analyses (Ortega and Lin, 2005; Sice, French et al, 2000).

The research presented here uses a combination of System Dynamics, control theory and non-linear analyses for modelling and exploring the behaviour of the centralised supply chain at Draeger Safety, Blyth, UK. The company is equipped with a sophisticated MRP environment for sharing information and enhancing communication between production and marketing. However, the performance of the supply chain is still perceived as problematic as it is not unusual that high inventories mount up and backlogs develop causing disruption as production is dealing with a crisis to overcome them. A centralised supply chain system was recently implemented with the purpose of diminishing costs and avoiding backlogs. This made Draeger's planning managers even more worried as it was difficult to predict what the consequences of centralised inventories would be for the manufacturing plant in Blyth. A team consisting of the planning department at Draeger Safety, UK and researchers from Northumbria University was formed to study the impact of different decision strategies on the behaviour of the supply chain. The purpose of the study was to model the material and information flows including the decision processes, provide explanation and insight into the link between decision strategies and inventory and backlog oscillations, and to develop a simple microworld that managers can use to conduct what-if scenarios and learn about the behaviour of the supply chain.

THE CENTRALISED SUPPLY CHAIN AT DRAEGER SAFETY, UK

Draeger Safety, UK is an international corporation, manufacturing breathing protection and gas detection equipment. It has a presence in

Figure 1. Schematic representation of the centralised supply chain proposed by Draeger Safety, UK



approximately 200 countries on all continents. Like many other industries Draeger Safety, UK makes use of a cascaded distribution system, with inventory holders at several different levels. In its simplest form (Figure 1), the central Hub, based in Germany, receives stock from the UK Factory and ships this to the main EU markets and to the Pacific and US Hubs. Each of the Hubs acts as a business unit and maintains a warehouse.

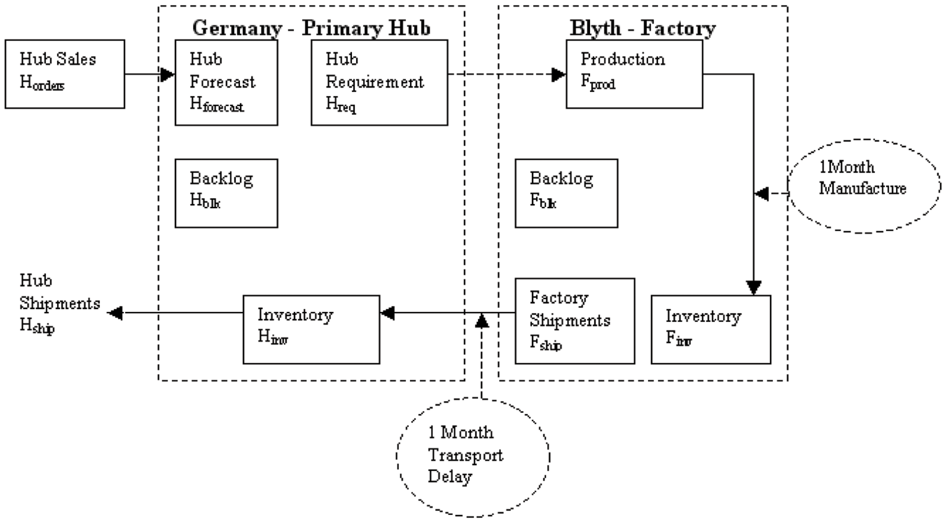
The Hubs receive orders from the Draeger Safety, UK sales agencies linked to them. The sales agencies do not keep inventory or receive shipments from the Hub. The role of the agencies is to receive orders from the customer and to make sales forecasts for their area. They then forward the orders and the forecasts to their respective Hub. The Hubs ship the goods directly to the customer and communicate with production units to help establish production requirements. The process is based on a 2 month forecast. That is, based on the current trends in orders and stock levels, the Hub forecasts its requirements for 2 months ahead. The factory then produces this in the current month so that it can be shipped to the Hub next month; in time to meet the expected demand.

The relationships that exist between the Central Hub and the Secondary Hubs are similar to those that exist between the Factory and the Central Hub. The research was conducted at the Draeger factory in Blyth. Thus, to better understand the operation of the supply chain, a simplified system dynamics model was constructed representing the relationships between the Factory and the Central Hub. The boundaries of the model were chosen based on discussions with Draeger Safety, UK staff. These determined that the model should include the customer, the production plant, the Hub in Germany and the relevant decision making processes. The structure of this portion of the supply chain is illustrated in Figure 2.

THE MODEL

The equations that are used to describe the inventories, backlogs and shipments are discrete (using a 1 month sample interval) and are based loosely on previous analysis of the Beer Game (Sterman 1989; Thomsen et al 1992). These equations are given in Appendix I.

Figure 2. Schematic diagram showing relationships between Factory and Central Hub. The figure represents a 'way it's meant to work' model of the supply chain at Draeger Safety, UK.



The management policies that are used to generate Hub orders and Factory production are based on an anchoring and adjustment heuristic (Sterman 1989). The HUB forecast forms the main driver for the ordering policy. Decision makers operate a form of adaptive forecast, which is formed from incoming orders based on a simple first-order exponential prediction. The mechanism underlying this forecast is characterised by an expression, which takes the form:

$$H_{forcast}(t+2) = (1 - \theta)\, H_{orders}(t) + \theta\, H_{forcast}(t+1) \quad (1)$$

Here, θ represents the rate at which the forecast is adapted. θ = 1, corresponds to a constant forecast, and θ = 0 describes a situation where the immediately previous value of HUB orders is used as an estimate of future demand.

In addition, decision makers were found to adjust orders above or below the expected demand in order to keep their inventory and supply line at the desired level. Thus, the overall ordering policy (HUB request & Factory production) may be characterised by the expressions:

$$H_{req}(t+2) = \max(0, \alpha\, (Q - H_{inv}(t) + H_{blk}(t)) - \alpha\, \beta\, (F_{blk}(t) + F_{ship}(t)) + H_{forecast}(t+2)); \quad (2)$$

$$F_{prod}(t) = \max(0, \alpha\, (Q - F_{inv}(t) + F_{blk}(t)) + H_{req}(t+2)); \quad (3)$$

where, α, (which lies in the range 0-1) defined the fraction of the discrepancy between the desired inventory Q and the effective inventory (Inventory – Backlog), that need to be ordered each month. β (which again lies in the range between 0 – 1) is the fraction of the supply line (requests or production already ordered but not yet received) that is taken into account. If β = 1, decision makers fully recognise shipments already in the pipeline and do not double order. If β = 0, requests (orders) are immediately forgotten and new requests to cover the same lack of inventory are placed in the next month.

Appendix II gives a Simulink block diagram representing the system equations.

MODEL ANALYSIS

An inspection of the model equations (Equations 14 to 19 and Appendix I) indicates that the primary nonlinearities present in the model are of a saturation form. Because of this, initial model analysis is based on a technique commonly used in control theory and employs a locally linear state space representation, in which all backlogs are assumed to be zero. Such an analysis will allow the soundness (or otherwise) of the basic management (feedback) structure to be assessed and will also provide an insight into the various transitions towards instability, without this information being lost in the added complications introduced by nonlinear behaviour.

Thus, if we define the general discrete state space representation as:

$$X(k) = A\,X(k\text{-}1) + B\,U(k) \tag{4}$$

$$Y(k) = C\,X(k)$$

The state vector X(k) as:

$$X(k) = \begin{bmatrix} H_{ship}(t) \\ H_{inv}(t) \\ H_{req}(t+2) \\ H_{forcast}(t+2) \\ F_{ship}(t) \\ F_{inv}(t) \\ F_{prod}(t) \end{bmatrix} \tag{5}$$

And, the input U(k) as:

$$U(k) = \begin{bmatrix} H_{orders}(t) \\ Q(t) \end{bmatrix} \tag{6}$$

Then, it is easily shown that the system matrices can be written as:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -\alpha_H & -\alpha_H\beta_H & , & -\alpha_H & 0 & 0 \\ 0 & 0 & 0 & , & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 1 \\ 0 & -\alpha_H & \alpha_F - \alpha_H\beta_H & , & -\alpha_H & -\alpha_F & -\alpha_F \end{bmatrix} \tag{7}$$

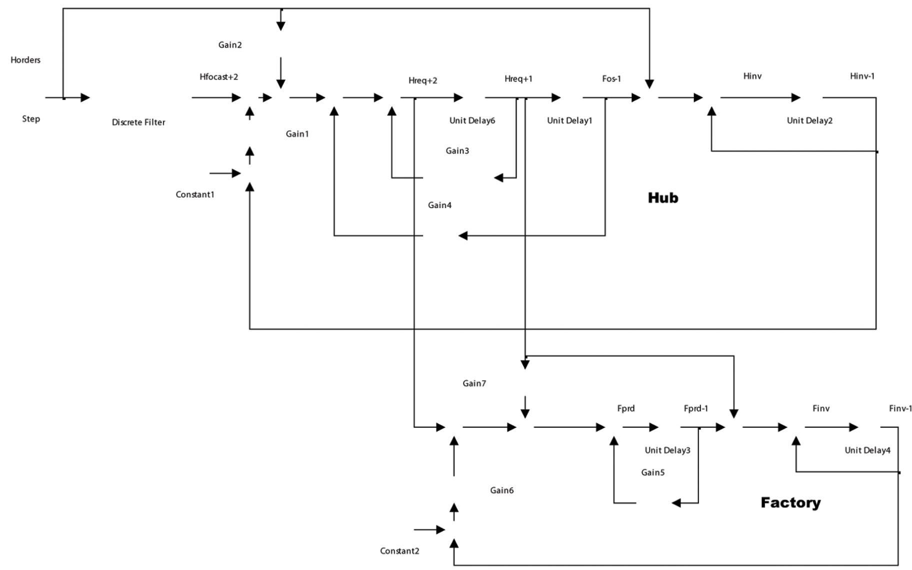
$$B = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 + \alpha_H - , & \alpha_H & 0 \\ 1 - , & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 + \alpha_H - , & \alpha_H & \alpha_F \end{bmatrix};$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{8}$$

and the system block diagram has the form shown in Figure 3.

In linear system analysis the relationship between behavioural modes and the location of the eigenvalues of A is well established. Consequently, it is clear that a plot of the trajectories of the eigenvalues in response to the variation of a specific parameter is a good indication of the influence that the parameter has on system behaviour. Inspection of Equation 7, indicates that the A matrix contains three variable parameters: α, β and θ. In studies of the Beer Game (Thomsen et al, 1992) it is generally accepted that for the majority of managers the value of θ tends to remain constant at 0.75. This is confirmed by finding at Draeger Safety, UK, where conversations with managers suggest that forecast is amended with a smoothing time of approximately 4 months (θ = 0.75). In addition, inspection of the Simulink block diagram (Figure 3 &Appendix II) indicates

Figure 3. Small signal block diagram representation of the unmodified model of the supply chain at Draeger Safety, UK



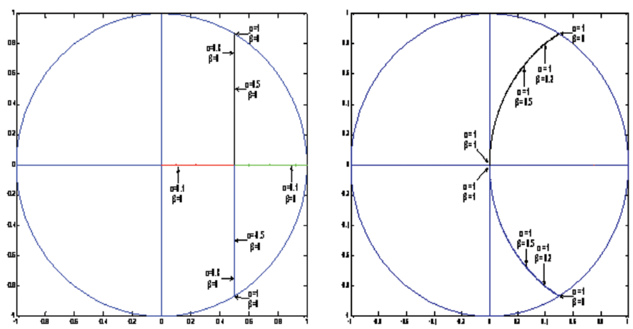
that θ is present only in the disturbance section of the model. Thus, θ only affects the dynamics of the feedforward (or disturbance) signal and can be considered to be outside of the feedback structure. α and β, on the other hand clearly lie inside the feedback loops and consequently will significantly influence the dynamic behaviour generated by the system itself. For these reasons, in the study presented here, θ will remain fixed at 0.75 and the analysis will concentrate only on variations in α and β.

THE INFLUENCE OF α AND β

Figure 4 shows a plot of the eigenvalue locus of the A matrix, calculated as the solution to det(zI - A) = 0 (where I is an identity matrix of dimension A): a) when β = 0 and α is varied from 0 to 1; and: b) when α = 1 and β is varied from 0 to 1. A quick inspection of the figure suggests that there is (at least) one pair of complex eigenvalues and that these intersect with the unit circle at a position 0.5 ± j0.866, for the condition α = 1, β = 0. Clearly, α has a destabilising influence, since the complex eigenvalue pair (Figure 4a) move toward the edge of the unit disc as α increase. β, on the other hand, is clearly stabilising, since the eigenvalues (Figure 4b) move toward the centre of the unit disc as β increases.

As can be seen in Figure 4(a) the condition α = 1, β = 0 generates a pair of complex eigenvalues that lie on the unit circle at a position 0.5 ± j0.866. These are responsible for the 'linear looking' (input dependant amplitude, 6 month period) oscillatory behaviour seen in Figure 5a. This behaviour represents one extreme of the management policy spectrum and is a typical consequence of a regime in which we seek to completely replenish any stock discrepancy at each time step. As α is decreased the oscillations become smaller. For example when α = 0.8, β = 0, the complex eigenvalue pair are located at 0.5 ± j0.745 and the damping ratio is approximately 0.2. The time response for this condition is shown in Figure 5b. As can be seen, the oscillations now decay to zero, but only after a considerable period; about 50 months. There is also a reduction in the maximum

Figure 4. Locus of eigenvalues of A plotted on the z-plane for θ = 0.75. a) β = 0 and α varied from 0 to 1. b) α = 1 and β varied from 0 to 1. Clearly, the complex pair migrates toward the edge of the unit circle as α increase, while they move toward the centre of the unit circle with increasing β.



(a) (b)

amplitude. Clearly (since β = 0), decision making is still based purely on current stock levels, therefore, the improvement in performance must come as a result of the relaxation in the requirement that stock must be maintained at a constant value. When α = 0.5, β = 0, the damping ratio increases to approximately 0.4. The corresponding time response is now relatively fast, with only approximately 25% overshoot and a 15 month settling time, as shown in Figure 5c. Finally, when α = 0.1, β = 0 all of the roots lie only on the real axis. Here the time response is purely exponential (Figure 5d), but once again the settling time has increased to around 50 months.

Figure 4 (b) shows a plot of the eigenvalue locus when α = 1 and β is varied from 0 to 1. Figure 6 shows the corresponding time response at β = 0.2 and 0.5. It is clear from the figures that β is basically stabilising, since the eigenvalues move toward the centre of the unit disc and the oscillations reduce as β is increased. It is clear from Figure 6 that as β increases not only do the oscillations die away more quickly, but also, the steady state level of Hub inventory decreases as a function of both β and the order level. Thus, as β increases, the gap between the desired (Q = 800) and actual Hub inventories increases. For example, in Figure 6b when β = 0.5, the Hub inventory settles at a value of 600 for the initial order level of 400/month, the inventory then reduces to a steady value of 400 when the order level is increased to 800/month. This effect is easily predicted from Equation 2 where the term $\beta x H_{forecast}$ clearly acts as disturbance on the stock correction term $Q - H_{inv}$ (inspection of Equations 2 & 3 shows that this will not be the case for the Factory in which the inventory will settle at Q).

To better understand the observed behaviour, we need to re-examine the structure of the system block diagram (Figure 3). Inspection of the figure reveals that the Factory and the Hub have 'isolated' feedback structures. Therefore, by considering their individual characteristic equations, these can be analysed independently.

Thus, for the condition β = 0 (depicted in Figure 4a), the Factory characteristic equation is:

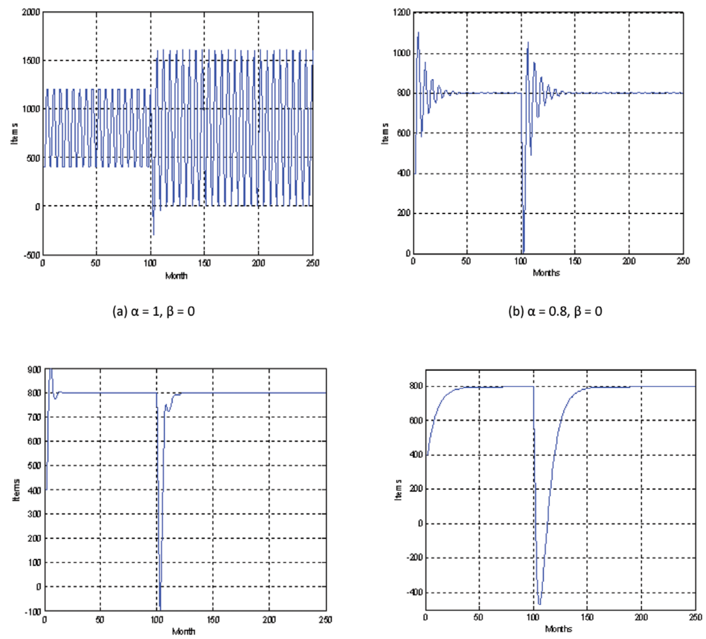
$$(z + \alpha)(z - 1) + \alpha = 0 \qquad (9)$$

This has two eigenvalues, one at z = 0 and a second, which is always real and which lies in the range z = 1 → 0 as α = 0 → 1.

The Hub characteristic equation is:

$$(z^2 + \alpha)(z - 1) + \alpha = 0 \qquad (10)$$

Figure 5. Hub effective inventory for θ = 0.75, α = 1, 0.8, 0.5, 0.1, β = 0 and Q = 800. The figure shows the response of the Hub effective inventory over a 250 month period. The initial Hub inventory is 400 and the Hub orders are held constant at 400/month for the first 100 months these are then increased to 800/month for the remainder of the simulation.



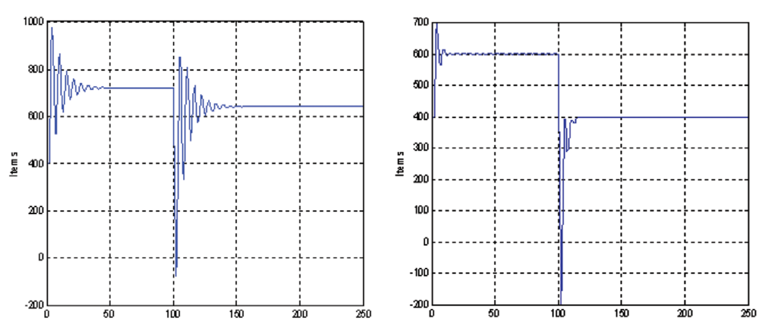
(a) α = 1, β = 0

(b) α = 0.8, β = 0

(c) α = 0.5, β = 0

(d) α = 0.1, β = 0

Figure 6. Hub effective inventory for θ = 0.75, α = 1, β = 0.2, 0.5 and Q = 800. The figure shows the response of the Hub effective inventory over a 250 month period. The initial Hub inventory is 400 and the Hub orders are held constant at400/month for the first 100 months these are then increased to 800/ month for the remainder of the simulation.



(a) α = 1, β = 0.2

(b) α = 1, β = 0.5

This has three eigenvalues. Again one of these is at $z = 0$, the other two form a second order pair that become complex when $\alpha > 0.25$. It is this pair that is clearly identified in Figure 4a. Moreover, it is the Hub's dynamics and not the Factory's that are the potential source of unstable behaviour. The Hub, potentially, becoming unstable for any value of $\alpha > 1$, whilst the Factory would be stable for any value of $\alpha < 2$.

THE EFFECT OF PRODUCTION LEAD-TIME

To better understand Factory/Hub interactions it is intended to investigate the effect of adding an additional two months lead-time into the factory manufacturing process; Figure 7. This additional delay brings the model into line with the classical Beer Game representation. However, unlike in the Beer Game, where the delay occurs naturally due to the nature of the brewing process, here, such an additional delay would tend to be the result of some unforeseen external influence: Perhaps a lack of material supply (due to industrial action by a supplier) or a prolonged upturn in business that outstrips the ability to produce. This form of analysis will allow the resilience of the management structure to be assessed and should provide some insight into the system's ability to recover from unexpected events, since in effect we are forcing the supply chain to operate in a regime where orders arrive 2 months behind expected need.

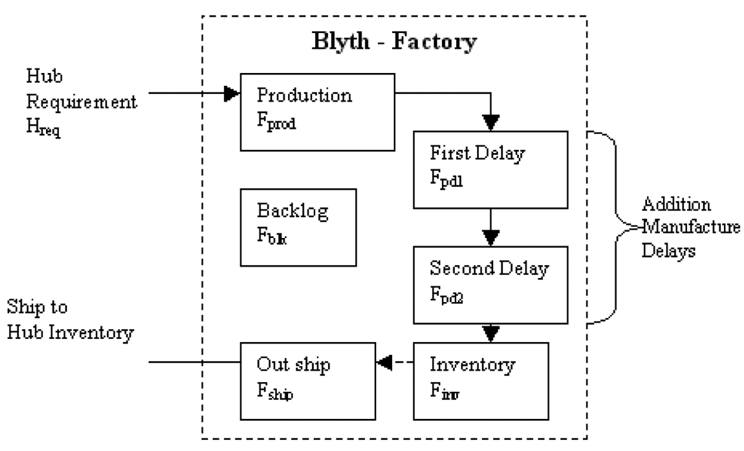
Since the model operates on a one month sample time, 2 additional factory states, F_{pd1} and F_{pd2}, are included to represent the additional manufacturing lead-time. The modified factory equations are given in Appendix III, and again it is relatively easy to show that the A matrix for this modified Factory/Hub arrangement is given by Equation 11.

The system block diagram has the form:

Figures 9 show eigenvalue plots for the modified A matrix for a similar range of values of α and β to those used to generate Figures 4.

The most obvious difference between Figures 9 and Figure 4 is the appearance of a second pair of potentially unstable, complex eigenvalues. On closer inspection, we can deduce that the original set of complex eigenvalues (seen in Figure 4 as the always stable complex pair) must be attributable only to the Hub, since their behaviour and locations are totally unaffected by the introduction of the two additional Factory states. Moreover, it must be the Factory, with its two additional delays, that is the source of the potentially unstable be-

Figure 7. Schematic diagram showing relationships between Factory and Central Hub. The figure represents a 'Production delay' model of the supply chain at Draeger Safety, UK.



Equation 11.

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\alpha_H & -\alpha_H\beta_H & \theta & -\alpha_H & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & -\alpha_H & \alpha_F - \alpha_H\beta_H & \theta & -\alpha_H & -\alpha_F & -\alpha_F\beta_F & -\alpha_F\beta_F & -\alpha_F \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

haviour. Figure 10 shows the time response of the modified model when θ = 0.75, α = 0.8, β = 0.1 and Q = 800. An inspection of Figure 9 shows that for these values there are a pair of complex eigenvalues (those attributable to the Factory) that lie outside the unit circle (the Hub complex eigenvalues of course still lying inside the unit circle); and it is these that are responsible for the observed oscillatory response.

Once again, to understand the source of this observed behaviour we need to re-examine the structure of the system block diagram; in this case, Figure 8. From the figure, it is clear that the Factory and the Hub still retain their 'isolated' feedback structures. Moreover, closer inspection reveals that the additional production delays only have an effect on the Factory 'loop'. Consequently, the Hub characteristic equation (Equation

Figure 8. Small signal block diagram representation of the model of the supply chain at Draeger Safety UK, including additional production delays

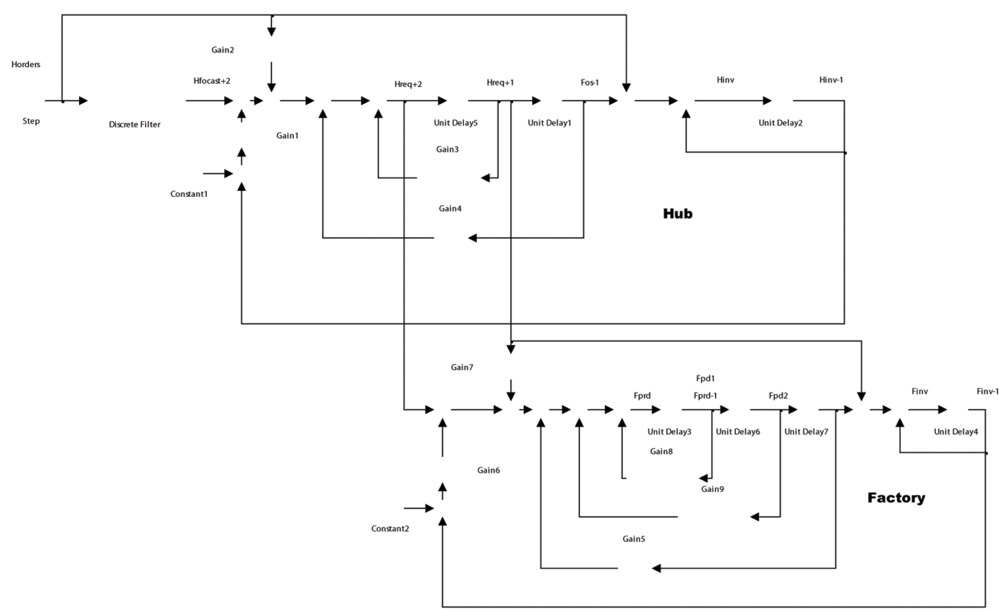
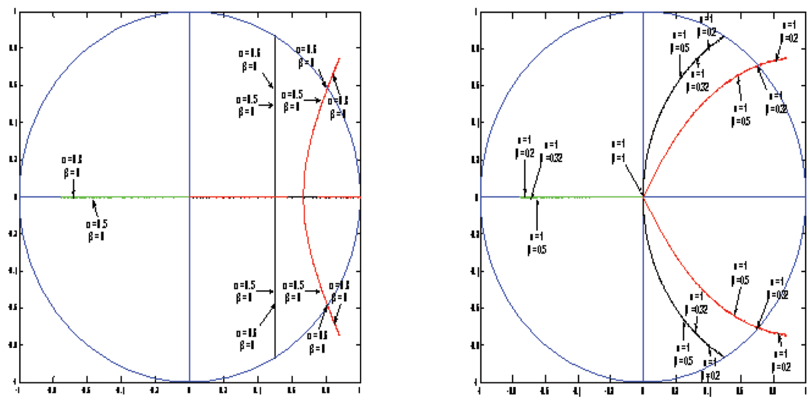


Figure 9. Locus of eigenvalues of A plotted on the z-plane for θ = 0.75. a) β = 0 and α varied from 0 to 1. b) α = 1 and β varied from 0 to 1



(a) (b)

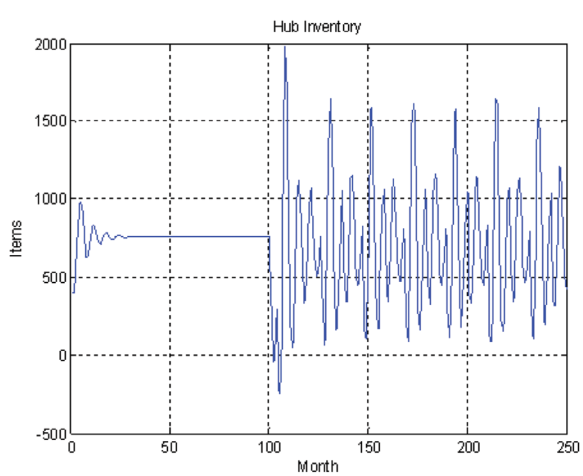
10) is unaffected by the additional production delays. Thus, the complex pair of eigenvalues previously identified as been associated with the Hub remains unaffected and are still clearly visible in Figure 9.

Turning our attention to the factory, the characteristic equation for the system with additional production delays, for the condition $\beta = 0$, is:

$$(z^3 + \alpha)(z - 1) + \alpha = 0 \qquad (12)$$

Clearly, Equation 12 has four eigenvalues. One of these is at $z = 0$, one exists on the negative real axis and lies in the range $z = 1 \rightarrow -0.755$ as $\alpha = 0 \rightarrow 1$ and the other two form the second complex pair seen in Figure 9. These become complex when $\alpha > 0.148$ and intersect with the unit circle

Figure 10. Hub effective inventory for θ = 0.75, α = 0.8, β = 0.1 and Q = 800. The figure shows the response of the Hub effective inventory over a 250 month period. The initial Hub inventory is 400 and the Hub orders are held constant at400/month for the first 100 months these are then increased to 800/ month for the remainder of the simulation.



at a position 0.809 ± j0.5878, for the condition α = 0.618. It is obvious, therefore, that it is now the Factory that has become the primary route to instability.

INFLUENCES OF 'SATURATION' EFFECTS

Since Figure 9 predicts an unstable eigenvalue for the condition, θ = 0.75, α = 0.8, β = 0.1 & Q = 800, the 'stable' initial transient seen for the first 100 months of the response in Figure 10 is somewhat puzzling. However, since we have suggested that the unstable eigenvalue pair can be related to the Factory section of the model (and not the Hub whose response is seen in Figure 10), then it is in the Factory's responses that we need to look for an answer.

Figure 11 shows the Factory effective inventory that corresponds to the Hub response shown in Figure 10. As can be seen the response is oscillatory over the whole timeframe. This form of response is of course the result of the large signal (saturation type) nonlinearities defined in Equations 17 to 20 limiting the growth of the small signal instability defined by the eigenvalues. What is interesting, however, is that prior to the increase in customer orders (from 400 to 800 at month 100) the Factory effective inventory is always positive and, therefore, the Hub demand can be met. After the increase in Hub orders the factory is no longer able to meet the required demand when the desired saftey stock (Q) is 800. Hence, the oscillations seen in the factory response, cascade down to the Hub. Figures 12 and 13, show the equivalent responses when Q is doubled to 1600.

As can be seen with Q = 1600 the Factory effective inventory remains positive over the whole response. Hence, when viewed from the 'isolation' of the Hub, the system appears perfectly stable and the Hub stock is able to settle to a steady value.

Figure 11. The Factory effective inventory response that corresponds to the Hub response seen in Figure 10. That is for θ = 0.75, α = 0.8, β = 0.1 & Q = 800. All initial values are set to 400 and customer orders are increased to 800 after 100 months.

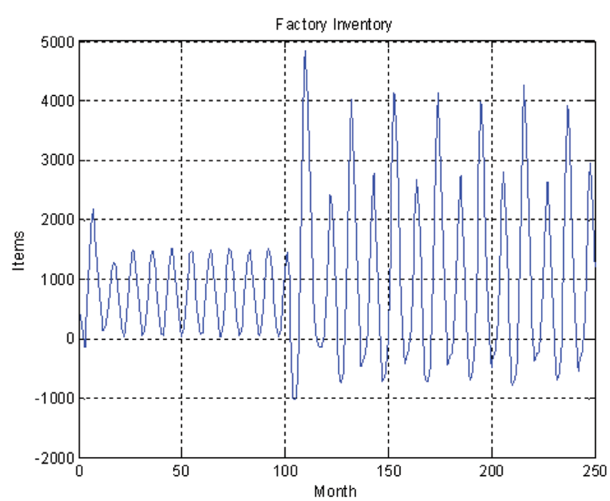
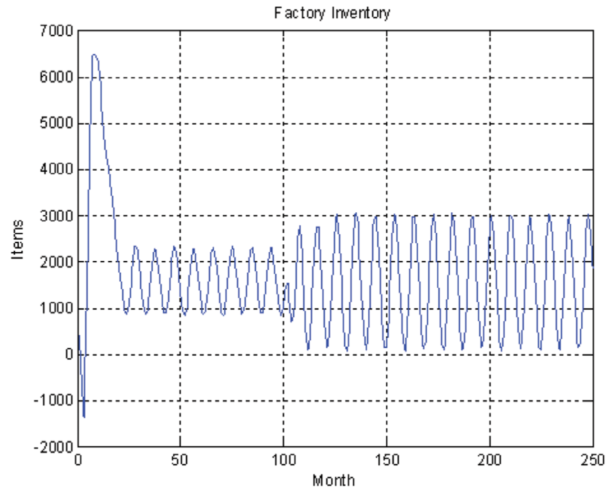


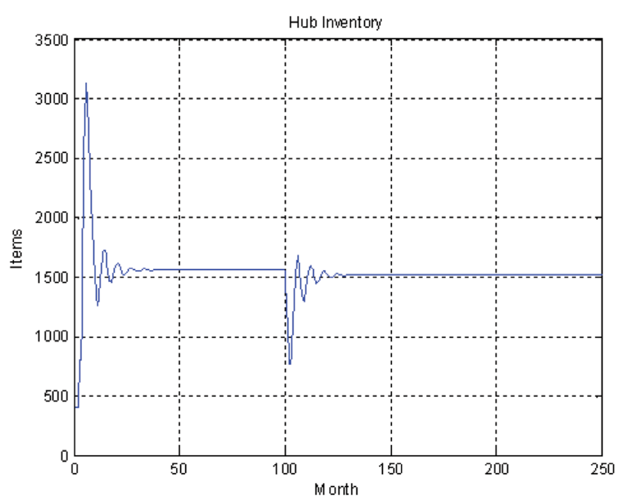
Figure 12. Factory effective inventory for θ = 0.75, α = 0.8, β = 0.1 & Q = 1600



DISCUSSION

In the model, management decisions are represented by the 3 corrective (negative) feedback loops. An initial estimate of Factory production is made based on the forecast for Hub sales in 2 months time. This estimate is then corrected (feedback 1) based on the discrepancy between the current Hub inventory and its desired level multiplied by a 'correction factor', α. This modified estimate is then re-corrected (feedback 2) to take into account any previously placed Hub orders (based on previous attempts to correct the Hub inventory discrepancy) that have not yet arrived and are thus 'still in the pipeline; in this case the multiplying factor is β. The final correction to Factory production (feedback 3), is made based on the discrepancy between current Factory inventory and its desired level, once again this is multiplied by α. In this form, the 'management

Figure 13. Hub effective inventory for θ = 0.75, α = 0.8, β = 0.1 & Q = 1600



policy' appears to be, primarily; one of stock control in the presence of external disturbances (actual product sales). Moreover, when operating close to their desired stock levels (and therefore without backlog) the feedback structures of Factory and Hub can be considered to be 'isolated' from one another. The feedback structure that is in place bares a strong similarity to the classical 'cascade control' structure, familiar to many engineers.

Within the context of this specified 'management policy' the two correction factors α and β, can be thought of as representing the psyche of the decision maker: $\alpha = 1$ and $\beta = 0$, represents one extreme; an aggressive manager thinking only within the confines of his area of control and not recognising the impact of previous decisions (no account of history). In the simulations this resulted in persistent oscillations (Figure 5a), generated purely by the decision making strategy and not as the result of structure of the supply-chain or some external influence (as is often claimed by managers with such tendencies). Such oscillations, while not impacting on customers (as inventories are in general maintained at positive levels) would require larger than necessary warehouse space. Decreasing α, and thus being less rigorous in terms of stock maintenance produces a stabilising effect; Figures 5b & 5c. If, however, α is too small (representing a very cautious manager) the response becomes very slow. This can lead to significant backlogs and the potential for customer dissatisfaction; Figure 5d. Clearly, there will be an optimum choice for α so that the response settles fairly quickly and remains backlog free.

$\beta > 0$, represents a manager that thinks outside of immediate area of control and takes some or full account of history. Taking into account the impact of previous decisions, has a stabilising effect; Figures 6a & 6b. As β increases the response becomes less oscillatory and tends to settle more quickly. $\alpha = 1$, $\beta = 1$ would seem to represent the ideal case (just in time, or deadbeat, management – with the eigenvalues lying at the origin of the z-plane). It is well known, however, that such control policies often prove unreliable, as they tend to be intolerant of unexpected events (such as unforeseen production hold-ups). Moreover, for the conditions simulated backlogs are significant and safety stock must be increased if customer satisfaction is to remain unaffected.

What is more interesting, however, in terms of the analysis presented in this paper is that it is possible to establish a relationship between the two major feedback loops (the Factory loop representing Factory management policies and the Hub loop representing Hub management policies) and potential system behaviours. For example, in the 'as is' arrangement (represented by Equations 9 & 10) the route toward instability is via the complex eigenvalue pair which are generated via the Hub management policy. More importantly, however, because the systems are 'isolated' poor management decisions in the Hub cannot be corrected by good decisions in the Factory. On the other hand, a good manager in the Hub may be able to mitigate poor management decisions in the Factory.

Figures 9 illustrate the influence of α and β, after the introduction of an additional 2 month manufacturing lead-time to simulate the effect of possible production hold-ups. Clearly, by comparing Figures 9 with Figures 4, we can see that α and β have much the same overall effect in this modified system. However, with the introduction of the extra states (additional lead-time), it is the Factory which now provides the primary route toward instability (via the second complex eigenvalue pair; Equation 12). In this situation the previous roles are now reversed and it is the Hub that can do little about poor management decisions in the Factory. Thus, good management and management policies are critical if significant problems are to be avoided (whereas in the case without the additional lead-time, not even the worst manager, if operating a consistent policy, could induce unstable behaviour). For example, Figures 11 and 12 show swings in factory inventory that are more that ten times greater that the disturbance that created them. These, if uncheck, would lead to a breakdown in customer confidence and the factory warehouse facilities grinding to a halt.

MANAGEMENT MICROWORLD

It is clear that the dynamics of the supply chain are intimately dependent on the decision making strategies of the managers. Faced with the prospect of a potentially unstable supply chain, Dreager commissioned the development of a virtual microworld (Senge, 1996) to allow managers to conduct what-if scenarios and to improve their understanding of how different decision strategies influence the performance of the supply chain. The first version of this Microworld was developed in Visual Basic and its interface is shown in Appendix IV. Managers are able to amend decision strategies by adjusting four decision parameters related to inventory, supply line, forecasting and level of desired stock. A graphical representation of the inventory and backlog oscillations allows a quick grasp of the link between decision strategy and the system's behaviour to be established. In addition, the relative cost of operating the supply chain is calculated as a sum of the factory and hub inventories over time and twice the hub's backlog:

$$\text{Costs} = F_{inv}(t) + H_{inv}(t) + 2\times H_{blk}(t) \qquad (13)$$

This is to take into account that backlogs are more expensive as they may lead to customers turning to other suppliers. A prototype of the Microworld is currently being used by the planning department. The planning team had already been involved in validating the model of the supply chain, and thus, had developed an understanding of how behaviour is linked with different decision strategies. Their feedback is valuable in further amending the interface to make it more user-friendly. The Microworld is to be used as a discussion tool for managers in the planning department and across departments to experiment and develop understanding what decision issues are important and what strategies work best.

CONCLUSION

The analysis presented here represents only the initial stages of an investigation into Draeger's supply chain. Moreover, in the results presented decision policies have been applied and maintained over an extended (and in real world term unrealistic) period of time. Never-the-less this analysis has found favour within Dreager Safety, UK and has proven itself to be a powerful decision support tool. It has become obvious that in our investigations so far we have only just scratched the surface of the rich variety of possible behaviours that can emerge. It is the intention, therefore, to continue this research and to further investigate:

- The influence of errors in the forecast.
- The effect of using independent decision strategies (having different values of α and β at the Factory and the Hub) at different levels in the supply chain.
- The impact of the secondary Hubs.
- Modifying the decision strategies (perhaps to a full state variable feedback form or production control in response to forecast) to take into account better information.

ACKNOWLEDGMENT

We would like to thank Mel Hedley, Head of Planning Department and Vince Smith, Operations Director, Draeger Safety, UK for their cooperation, good will and valuable contribution.

REFERENCES

Ackermans, H., & Dellaert, N. (2005). The Rediscovery of Industrial Dynamics: the contribution of system dynamics to supply chain management in a dynamic and fragmented world. *System Dynamics Review*, *21*(3), 173–189. doi:10.1002/sdr.317

Beer, S. (1979). *The Heart of Enterprise. New York*. Chichester: John Wiley & Sons.

Beer, S. (1994). *Decision and Control. New York*. Chichester: John Wiley & Sons.

Chang, Y., Makastsoris, H. (2001). Supply chain modelling using simulation. *International Journal of Simulating, 1.*

Cox, A., Watson, G., Lonsdale, C., & Sanderson, J. (2004). Managing Appropriately in Power Regimes: Relationship and Performance Management in 12 Supply Chain Cases. *Supply Chain Management: an International Journal*, *9*(5), 357–371. doi:10.1108/13598540410560748

Forrester, J. W. (1961). *Industrial Dynamics*. Portland, OR: Productivity Press.

Fowler, A. (1999). Feedback and Feedforward as systemic frameworks for Operations Control [MCB University Press]. *International Journal of Operations & Production Management*, 182–204. doi:10.1108/01443579910247428

Hong-Minh, S. M., Disney, S. M., & Naim, M. M. (2000). The dynamics of emergency transhipment supply chains. *International Journal of Physical Distribution and Logistics Management*, *30*(9), 788–815. doi:10.1108/09600030010351471

Li, H., & Li, L. (2005). Integrating Systems Concepts into Manufacturing Information Systems. *System Dynamics Review*, *21*(3), 135–147.

Lyneis, J. (2005). System Dynamics for Market Forecasting and Structural Analysis. *System Dynamics Review*, *21*(3), 5–17.

Marquez, A. C. Bianchi, Carmine, Gupta, Jatinder, N.D. (2004). Operational and financial effectiveness of e-collaboration tools in supply chain integration. *European Jounal of Operational Research, 159(*2).

Maturana, H., & Varela, F. J. (1987). *The Tree of Knowledge*. Boston, MA: New Science Library.

Morecroft, J. D. W. (1985). The Feedback View of Business Policy and Strategy. *System Dynamics Review*, *1*, 4–19. doi:10.1002/sdr.4260010103

Morecroft, J. D. W. (2007). *Strategic Modelling and Business Dynamics: A Feedback Systems Approach*. New York: Wiley.

Morecroft, J. D. W., & Sterman, J. D. (1994). *Modelling for Learning Organizations, System Dynamics Series*. Portland, OR: Productivity Press.

Mosekilde, E. (2002). *Topics in Nonlinear Dynamics – Applications to Physics, Biology and Economic Systems*, World Scientific, Singapour, 2002.

Naylor, J., Naim, M., & Berry, D. (1999). Leagility: integrating the lean and agile manufacturing paradigms in the total supply chain. *International Journal of Production Economics*, *62*, 107–118. doi:10.1016/S0925-5273(98)00223-0

Ortega, M., & Lin, L. (2004). Control theory applications to the production-inventory problem: a review. *International Journal of Production Research*, *42*(11), 2303–2322. doi:10.1080/00207540410001666260

Senge, P. M. (1984). *System Dynamics as an Artefact for the Systems Age. Dynamica, 10*. Part II.

Senge, P. M., (1996). *Systems Thinking*, Executive Excellence, January.

Sice, P. V., & French I.G., (1998). Control of Chaos in Dynamic Decision Making. *International Journal of Knowledge Transfer*, London, 49-55.

Sice, P. V., & French, I. G. (2000). Using system dynamics to analyse interactions in duopoly competition. *System Dynamics Review*, *16*(2), 113–133. doi:10.1002/1099-1727(200022)16:2<113::AID-SDR189>3.0.CO;2-G

Sterman, J. D. (2000). *Business Dynamics, Systems Thinking and Modeling for a Complex World*. New York: McGraw Hill.

Terzi, S., & Cavalieri, S. (2004). Simulation in the supply chain context: A survey. *Computers in Industry*, *53*, 3–16. doi:10.1016/S0166-3615(03)00104-0

Thomsen, J. S., Mosekilde, E., & Sterman, J. (1992). Hyperchaotic Phenomena in Dynamic Decision Making. *SAMS*, *9*, 149.

Van der Zee, D. J., & Van der Vorst, J. G. A. J. (2005). A modelling Framework for Supply Chain Simulation: Opportunities for Improved Decision Making. *Decision Sciences*, *36*(1). doi:10.1111/j.1540-5915.2005.00066.x

Venix, J. (1994). *Group Model Building: Facilitating Team Learning Using System Dynamics*. New York: Wiley.

Yusuf, Y. Y., Gunasekaran, A., Adeleye, E. O., & Sivayoganathan, K. (2004). Agile supply chain capabilities: Determinants of competitive objectives. *European Journal of Operational Research*, *159*(2). doi:10.1016/j.ejor.2003.08.022

APPENDIX A

Inventories in both the Hub and the Factory are updated by adding incoming shipments and subtracting outgoing shipments.

$$H_{inv}(t) = \max(0, H_{inv}(t-1) + F_{ship}(t-1) - H_{ship}(t)) \quad (14)$$

$$F_{inv}(t) = \max(0, F_{inv}(t-1) + F_{prod}(t-1) - F_{ship}(t)) \quad (15)$$

To the extent that inventory plus incoming shipments are sufficient, outgoing shipments are existing backlog plus incoming orders. Otherwise, outgoing shipments are incoming shipments plus inventory, and the new inventory is empty.

$$H_{ship}(t) = \min(H_{orders}(t) + H_{blk}(t-1), H_{inv}(t-1) + F_{ship}(t-1)); \quad (16)$$

$$F_{ship}(t) = \min(H_{req}(t+1) + F_{blk}(t-1), F_{inv}(t-1) + F_{prod}(t-1)); \quad (17)$$

In a similar way, backlogs are updated by adding incoming orders and subtracting outgoing shipments. If incoming orders plus backlogs are completely covered by incoming shipments plus existing inventory the new backlog is empty.

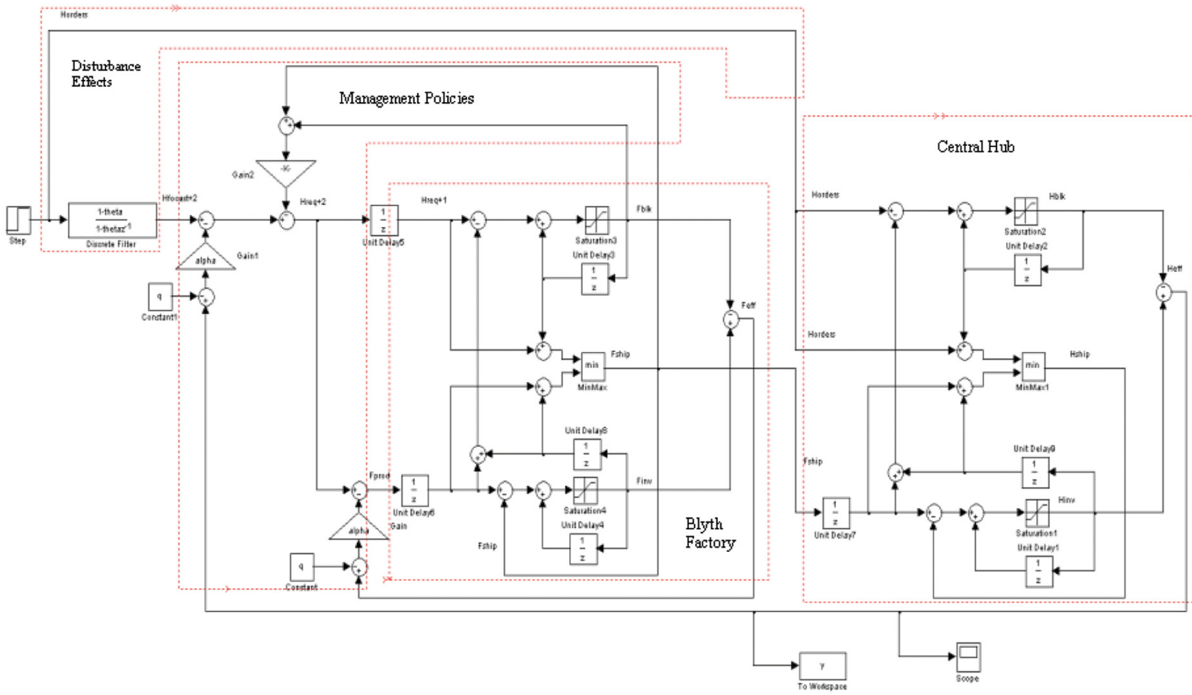
$$H_{blk}(t) = \max(0, H_{blk}(t-1) + H_{orders}(t) - (H_{inv}(t-1) + F_{ship}(t-1)); \quad (18)$$

$$F_{blk}(t) = \max(0, F_{blk}(t-1) + H_{req}(t+1) - (F_{inv}(t-1) + F_{prod}(t-1)); \quad (19)$$

APPENDIX B

Simulink diagram of the Factory - Hub model based on equations. Figure 14 shows a Simulink representation of the model. Inspection of the figure shows the series nature of the Factory-Hub arrangement and the similarity of the modelling relationships used to represent both the Factory and the Hub. It is also easy to identify the classic ‘cascaded control’ feedback structure, which represents the decision making process governing Factory production.

Figure 14.



APPENDIX C

Factory equations including the additional lead-time states:

$$F_{inv}(t) = \max(0, F_{inv}(t\text{-}1) + F_{pd2}(t\text{-}1) - F_{ship}(t)) \tag{20}$$

$$F_{ship}(t) = \min(H_{req}(t{+}1) + F_{blk}(t\text{-}1), F_{inv}(t\text{-}1) + F_{pd2}(t\text{-}1)) \tag{21}$$

$$F_{blk}(t) = \max(0, F_{blk}(t\text{-}1) + H_{req}(t{+}1) - (F_{inv}(t\text{-}1) + F_{pd2}(t\text{-}1)) \tag{22}$$

$$F_{prod}(t) = \max(0, \alpha\ (Q - F_{inv}(t) + F_{blk}(t))$$

$$+ H_{req}(t{+}2) - \alpha\ \beta\ (F_{pd1}(t) + F_{pd2}(t))) \tag{23}$$

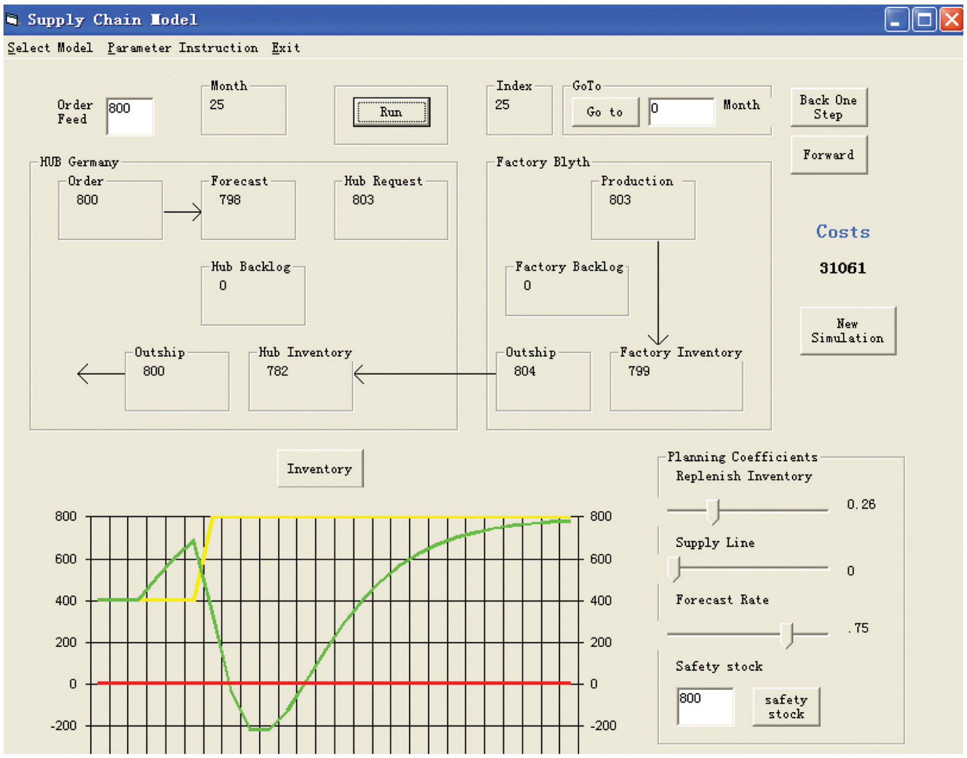
$$F_{pd1}(t) = F_{prod}(t\text{-}1) \tag{24}$$

$$F_{pd2}(t) = F_{pd1}(t\text{-}1) \tag{25}$$

APPENDIX D

Visual Basic Microworld interface

Figure 15.



This work was previously published in Supply Chain Optimization, Management and Integration: Emerging Applications, edited by John Wang, pp. 33-53, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 88
The Drivers for a Sustainable Chemical Manufacturing Industry

George M. Hall
University of Central Lancashire, UK

Joe Howe
University of Central Lancashire, UK

ABSTRACT

This chapter describes the current situation of the chemical manufacturing industry and looks to the future demands on the sector such as: for sustainability, the advent of new bio-based feedstocks for improved energy management and the implications of these demands on the sector. These implications include definitions of sustainability criteria for the chemical manufacturing industry and the need for transparent reporting following the Triple Bottom Line approach. The important role of chemical (or more generally, process) engineers in delivering bio-based sustainable solutions is emphasised, but this also suggests that a new way of thinking about the discipline is required. Indeed, there are arguments that the demand for a sustainable chemical manufacturing sector could bring about the next paradigm shift in the discipline with concomitant education implications.

INTRODUCTION

The Chemical Manufacturing Industry (CMI) can be multinational in operation and varied in scale whilst producing the products which underpin other sectors such as: health provision; clothing; housing and shelter; food and nutrition; entertainment and leisure; transport and tourism—the very fabric of human life. The variety of activity in the sector can be said to demand that, "Process Engineering," is a more correct term to encompass industries as widely different as food processing, minerals processing, pharmaceutical manufacture and petroleum refining. However, there is commonality in their activity where heat and mass transfer, transport phenomena, reaction kinetics, modelling and fluid mechanics are all applied in unit operations albeit to widely varying materials

DOI: 10.4018/978-1-4666-1945-6.ch088

and situations. For this reason in this chapter we have used the term CMI to cover this spectrum of activity and the general approaches which we describe can be applied equally well to all in our opinion.

The products of the chemical industry are derived from inorganic, synthetic organic and biological sources which have been manipulated by a range of process operations which include classical physico-chemical processes such as synthesis, distillation, precipitation, filtration, solvent extraction and crystallisation techniques which have been supplemented by fermentation processes, the application of industrial enzymes and the genetic manipulation of microorganisms in modern biotechnology. These technologies have made a vast number of products available to promote social development and economic growth and prosperity. At the same time the chemical industry has been accused of overexploitation of natural resources, air, water and land pollution (such as by oil production in the Niger River delta, opencast mining on fragile Pacific islands and the catastrophic Deepwater Horizon oil spill in the Gulf of Mexico) and creating social problems associated with rapid industrialisation and the invasion of vulnerable societies and environments for commercial gain (witness the toxic release from the Union Carbide plant in Bhopal, India, in 1984). When Industrial Revolutions have taken place starting from Europe, North America and then into Latin America, Asia and Africa the ills of industrialisation were ignored initially, as the price to pay for progress although eventually the development of organised labour and government structures have improved the situation in many cases. Whilst accepting this historical context (Coley & Wilmot, 2000) the future development of the CMI must nowadays be viewed in the light of the concept of sustainability (Garcia-Serna, Perez-Barrigon, & Cocero, 2007) which means tackling the issues of the change from fossil fuel to bio-based raw materials and energy conservation and management (Hall & Howe, 2010).

This chapter will describe the criteria by which the CMI is judged and how sustainability is reported by the industry. It will then describe the new technologies impacting on process engineers in a sustainable CMI and finally go on to discuss whether the sustainable approach will demand a new type of process engineer and the educational implications of this demand.

BACKGROUND

Defining Sustainability

Sustainability, or Sustainable Development (SD), is a discipline that has long been of interest to scientists, technologists, politicians, and business alike. Thomas Malthus in his, *Essay on the Principle of Population*, of 1798, proposed that the power of population increasing in a geometric ratio would outstrip the power of the earth to sustain mankind increasing in an arithmetic ratio and thus a link was made between population and sustainability. However, the advent of the concept of sustainability can be traced back in modern times to the 1970's but the UN Commission on Environment and Development (the so-called Brundtland Report) in the 1980's was a major point in defining the topic (World Commission on Environment and Development, 1987). The Commission was originally instructed to investigate the issues of global inequality, resource distribution and global population impacts and recommend solutions to these issues. Their oft-quoted report definition of sustainable development is, "Development which meets the needs of the present without compromising the ability of future generations to meet their needs."

The report linked economic development with social and environmental concerns for the first time and a balance of economics, social justice and environmental protection was proposed if sustainable development was to happen - this became known as the Triple Bottom Line. The

Brundtland Commission was the starting point for the many UN-sponsored events and initiatives such as the Rio Summit in 1992 on environment and development and the Kyoto Protocol (1997) on greenhouse gas emissions and onwards through a series of negotiations leading up to the inconclusive COP 15 meeting in Copenhagen in 2009 which is still being debated today.

Early implementation of these ideas in industry linked long-term sustainability with competitiveness where obvious links existed between, say, energy efficiency and savings which would benefit the traditional (financial) bottom line and the TBL (Florida, 1996; Judge & Douglas, 1998). Other studies have linked sustainability with organisational capability leading to increased competitiveness (Aragon-Correa & Sharma, 2003) and the promotion of sustainability into corporate strategy (McGee, 1998). More recently there has been the introduction of sustainability considerations at the operational level which would assess the viability of new projects, the introduction of new technology and even overall company structure against the background of sustainability (Labuschagne, Brent, & Van Ercka, 2005).

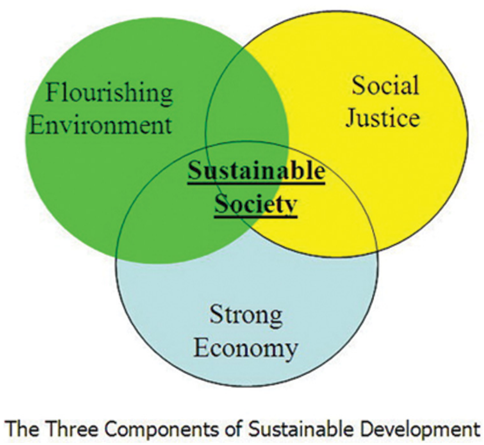
However, as mentioned above, the CMI is a very varied sector covering a range of raw materials and processing operations; different scales and global impact of operation and differing environmental impacts as a consequence of its activities. As a result we should ask ourselves: is there a coherent approach to sustainability in the sector and are the current chemical engineers being educated in sustainability appropriate to the needs of a sustainable industry? Batterham (2006) argued that process engineers would play a significant role in achieving sustainability goals through their understanding of the molecular and micro levels and their integration into macro systems together with their abilities in systems analysis, modelling and process balances. Furthermore, addressing these different levels would also make process engineering practice assess the sustainability agenda as a whole - chemical engineers must change the way they practice and embrace sustainability concepts and implement them.

The Triple Bottom Line and Corporate Social Responsibility

The Triple Bottom Line (TBL) (Figure 1) has emerged as a popular conceptualization and reporting vehicle for articulating corporate social, environmental, and economic performance and is arguably the most accepted impact review model within 'green' or 'socially responsible' businesses since the 1990s (Brown, Dillard, & Scott Marshall, 2006). It has been embraced by business and commerce to legitimise their activities in the eyes of the general public, government and environmental watchdogs. The concept of Corporate Social Responsibility (CSR) has been developed to allow businesses to measure and promote their impact on the three pillars of the TBL and several definitions of CSR exist (Carroll, 1999) often reflecting the viewpoint of the definers. However, they usually contain concepts of the integration of social and environmental concerns alongside economics and the interaction of the business with its employees, the local community and society at large. The term CSR is often associated with activities of conventional business rather than those social enterprises or non-profit-making enterprises where social and environmental concerns are paramount in their philosophy. There is of course some scepticism over the sincerity, practicality and even the legitimacy of this approach given the wide spectrum of business scale, operating styles, cultural attitudes and the strength of national governments to monitor and enforce a TBL model. Some examples of CSR activity include:

- corporate philanthropy and sponsorship although this can be seen as pure public relations
- local community investment and support for worthy causes although again there are

Figure 1. The triple bottom line



concerns over the influence of business in the public arena and which worthy causes are chosen

- stakeholder engagement brings dialogue between the business and all those affected by its activities as long as the definition of a stakeholder is sufficiently wide
- social and environmental reporting being given equal weight to financial reporting. Such reporting can be used to hold a business to account especially if done in conjunction with corporate codes of conduct which lay out the company, "values," and standards of behaviour. This aspect of CSR will be discussed later as it is seen as an essential tool to judge progress towards sustainable business operation.

SUSTAINABILITY ASSESSMENT

The sustainability of any enterprise or activity can be described by two Carbon-management concepts, Carbon Footprinting and Carbon Labelling, which are popular indicators to the general public that their purchasing is ethical while a third concept, Life Cycle Assessment (LCA) also can be used to investigate the sustainability impact of the CMI. Increasingly, these concepts are being applied across the, "Supply Chain," which brings into focus the activities of the companies which supply the central operation under assessment (the CMI here) and also those companies they sell to or who distribute and dispose of their products. The development of these tools to describe the impact of industry activity on the environment is essential if the issues involved are to be argued in reasoned rather than emotive terms. They also give credence to claims by governments, industry and interest groups which can be trusted by the general public.

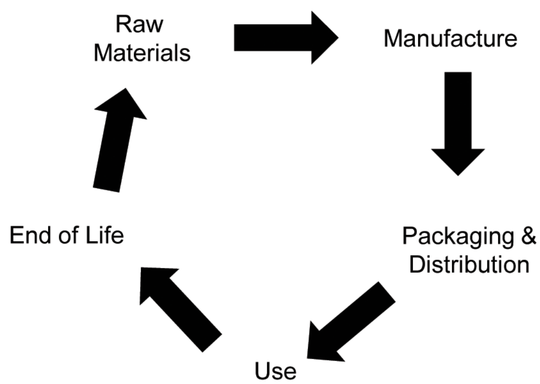
Life Cycle Assessment

Life Cycle Assessment is the investigation and evaluation of all the environmental impacts of a given product, process or service. Concepts such as Carbon Footprinting and Ecolabelling (see below) rely on LCA as it provides a methodology which has some international standing and uniformity. An early definition of LCA was, "the process of evaluating the effects that a product has on the environment over the entire period of its life cycle…extraction and processing; manufacture; transport and distribution; use, reuse, and maintenance; recycling and final disposal" (United Nations Environment Programme, 1996).

There are variants on the basic LCA definition, a full LCA is sometimes called, "cradle-to-grave," including all activity from raw material sourcing to disposal of all components after use (Figure 2).

Where a product generates no waste because all materials are recycled or converted to biologically-safe materials the term, "cradle-to-cradle," is used. The LCA can be used for specific elements of a product stream such as energy, water, packaging and raw materials. The core elements of LCA are available in several software packages which lead the user through the phases, provide generic categories and conversion factors and impact assessment models. Variants for specific applications and sectors abound and international cooperation

Figure 2. Life cycle assessment: cradle-to-grave



has lead to greater uniformity and consolidation of methodologies. Case studies for a range of manufacturing processes and products are given in Table 1 which also indicates the issues which were raised by each case study.

The LCA is a powerful tool in the analysis of commercial activity, including the CMI, so will be described in some detail here in order that interested parties can see the relevance to their own situation. The description of an LCA process here is based on that of the International Standards Organisation 14040 Series which is a generally recognised (International Standards Organisation, 2006).

As shown in Figure 3 an LCA will normally be divided into four activities:

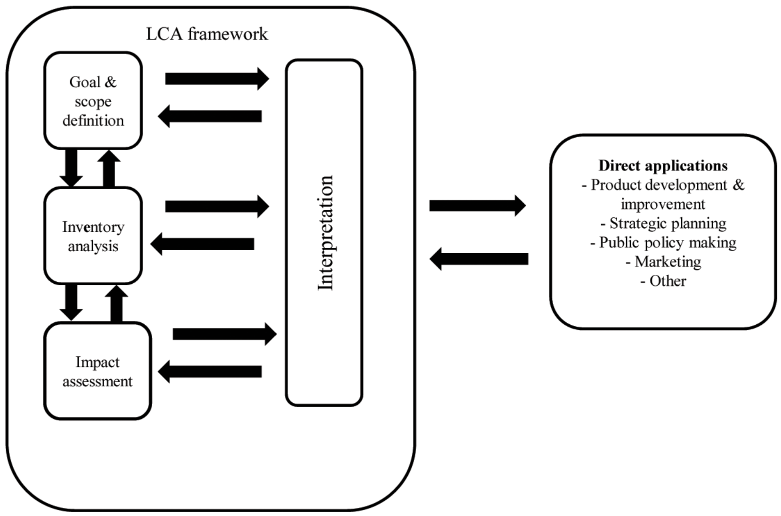
***Phase 1: Goal and Scope*:** the goal decides whether all or specific aspects (e.g. energy) of the operation are included. The system boundaries are set by the scope definition and can be divided into four areas: (1) pre-manufacture, (2) manufacture, (3) packaging and distribution, and (4) use and end of life. Consideration of all areas would be, "cradle-to-grave," as mentioned above, but can be more limited such as, "cradle-to-factory gate," which would only include the first two areas. This phase also decides the purpose of the LCA and for whom it is being done as this will affect the data collected and its conversion to meaningful units. If a processing plant produces more than one product (co-products) then there must be an allocation of impact between them. The simplest approach, system allocation, is made on the basis of the mass or the economic value of the product but does discern differences in process operations which lead to the co-products – not all process operations have equal impact. An alternative approach (preferred by the ISO 14040 series) is system expansion (Weidema, 2003).

***Phase 2: Inventory analysis*:** is a data collection phase and includes all inputs (e.g. energy and raw material), outputs (products) and emissions (to air, water, soil, and solids) or those selected

Table 1. Life cycle assessment case studies in the manufacturing industries

Products/Process	Scope	Issues	Reference
Cocoa	Cradle-to-grave	Missing/unreliable data	Ntiamoah and Afrane (2008)
Tomato ketchup	Cradle-to-grave	Capital equipment inputs Wholesale & retail stages omitted	Andersson, Ohlsson and Olsson (1998)
Paper pulp	Cradle-to-factory gate	Enzymatic processing to reduce environmental impacts	Skals, Krabek, Nielsen & Wenzel (2008)
Wood-based products	Cradle-to-grave	Waste recycling or energy by combustion	Jungmeier, Werner, Jarnehammar, Hohenthal and Richter (2002)
Personal computer	Cradle-to-grave (cradle?)	Recycling rates Energy consumption in use	Choi, Shin, Lee & Hur (2006)
Fine chemicals (pharmaceuticals)	Cradle-to-factory gate	Environmental impact and molecule complexity Energy consumption	Wernet, Conradt, Isenring, Jimenez-Gonzalez & Hungerbuhler (2010)

Figure 3. Stages in an LCA



for inclusion. Accurate, relevant information is essential and must be available or derived from secondary data such as utility bills for water, gas and electricity. This activity is the most time consuming and challenging if a company has not attempted any such exercise before. A production system can be broken down into unit processes, batch or annual production, whichever best defines the system in meaningful numbers, called the functional unit.

***Phase 3: Life Cycle Impact Analysis*:** in this activity the inventory analysis information is processed and first of all assigned to an environmental impact category with appropriate units which may conform to systems such as ISO 14000 Series or be process-specific. Six common environmental impact categories are:

- ***global warming*** where the main contributor is combustion of fossil fuels expressed as carbon dioxide equivalents
- ***acidification*** which affects waters, forests and caused mainly by combustion for electricity, heating and transport and expressed as sulphur dioxide equivalents
- e***utrophication*** which leads to algal blooms, oxygen depletion and fish deaths is caused mainly by fertiliser nitrogen run off expressed as nitrate equivalents
- ***ozone depletion*** caused by man-made halocarbons
- ***land use*** in the production of products and expressed as by unit area per year
- ***photochemical*** smog from volatile organic compounds produced from unburnt petrol and diesel and organic solvents causes respiratory problems and reduces agriculture yields—expressed as ethane equivalents.

These categories are not exclusive and for certain applications the energy, water and effluent categories can be simplified from those above or made process-specific (Azapagic, Clift, & Predan, 2004). Once categorised any emissions should be converted to the reference units for that category using equivalence factors. A sensitivity check and normalisation process will determine the accuracy and relevance of the data collected. Finally, the inventory data can be weighted in terms of the most important environmental impact. The weighting criteria are, again, areas of debate and can be based on: the judgement of a panel of experts; financial considerations and targets set by the company or government edict.

***Phase 4: Life Cycle Interpretation*:** the results of the impact analysis are compared with the original goal and scope of the project and judged, somewhat subjectively, against them. This analysis need not be left until the end but can take place continuously to ensure that the LCA is really achieving the goals and the scope is correct. This iterative approach to the interpretation of data will allow incremental improvements and/or changes to the goals and scope as necessary. The final interpretations should indicate the completeness of the data, the appropriateness of the analysis and reach conclusions and lead to recommendations for process improvement.

The LCA as described above has been applied successfully to a wide range of activities and processes and so should be applicable to the wide ranging CMI. It has evolved to cover environmental, cost and social aspects and as these different versions all have an importance for the CMI and a brief description of each is appropriate here. The early development of LCA can be traced back to the late 1960s – early 1970s (Boustead, 1996) and was a general environmental impact assessment tool (E-LCA now) which evolved by introducing separate categories to reflect new uses. An understanding these impact factors led to Life Cycle Costing (LCC) to assess the financial, time and resource costs of the above impact factors, calculating, for example, the amount of pollution caused distribution methods.

An E-LCA rarely included elements of human or social assessment directly but they can have some element of human welfare included within them. Azapagic *et al.* (2004) noted that the impacts of persistent chemicals on air, land and water lead to human, animal and eco-system damage. However, including societal elements into LCAs is difficult as they are not necessarily measurable physical attributes and therefore difficult to quantify. So, finally, it should be noted that the concept of the Social Life Cycle Assessment (S-LCA) has been developed recently to complement the traditional E-LCA and in doing so give a fuller assessment of the sustainability of goods and services (Benoit, et al., 2010). Guidelines for the S-LCA of processes and products have been drawn up (Benoit & Mazijn, 2009) putting the process into the context of E-LCA. An important goal for the S-LCA has been proclaimed as the improvement of the social conditions of stakeholders in the process/product covered by the S-LCA (Jorgensen, Finkbeiner, Jorgensen, & Hauschild, 2010). Stakeholders have been defined (Benoit & Mazijn, 2009) as: workers, consumers, the local community and society at large, consumers (all stages) and value chain actors and, within these, there are many subjective impact categories such as child labour, fair salary, equal opportunity, migration, secure living conditions and technology development. Thus S-LCA claims to assess the social and socio-economic impacts of all life cycle stages and is a systematic process using best available science to collect best available data on, and report about, social impacts (positive and negative) in product life cycles from extraction to final disposal (Benoit, et al., 2010). The feasibility of integrating social aspects into LCAs seems challenging but it could prove to be a rewarding exercise. Griesshammer *et al.* (2006) concluded that there are many hurdles to overcome for S-LCA to be comprehensive and needed:

- to improve the data situation (qualitative, semi-quantitative and/or fully quantitative)
- carry out more case studies
- establish an accepted social indicator list (including impact categories)
- a connection with CSR needs to be realised and implemented.

Objective indicators measure impacts (i.e. wages, working hours) whereas subjective indicators focus on the experiences or feelings of the relevant stakeholder (i.e. assessing the wealth of society as a broad notion) (Jorgensen, et al., 2010). It is difficult to assess how the concept will achieve these tasks and truly measure the needs of creatures as complex as humans and their complex social environments. The two most pressing research needs are for case studies and for the development of methodological sheets for inventory analysis. It is obvious that the S-LCA could become a useful tool to promote the activities of the CMI or a stick to beat it. Nevertheless, it will become incumbent on the CMI to get to grips with these new assessments of its activities.

Carbon Footprinting and Labelling

Carbon Footprinting (CF) is a measure of the Greenhouse Gases (GHG) associated with the manufacture of products taking into account: raw material sourcing, processing, packaging and distribution and waste treatment. Emissions are described as, "direct," when associated with the main process or activity and, "indirect," when associated with upstream and downstream activities. The term GHG can be defined as carbon dioxide only, or include other gases, particularly methane (with a GHG effect over 20 times that of carbon dioxide), as carbon dioxide equivalents. In some cases it is difficult to convert an activity or process into GHG units and conversion factors are not uniform. Hence several definitions of CF have been proposed some more rigorous than others, as described by Wiedmann and Minx (2007), who proposed the following definition of the term, "The carbon footprint is a measure of the exclusive total carbon dioxide emissions that is directly and indirectly caused by an activity or is accumulated over the life stages of a product."

As mentioned above, the size of the CF can vary depending on how many core and peripheral activities relating to product are included e.g. the activities of suppliers to the principal manufacturer and even the GHG associated with the consumer travelling to buy the goods. Again, this is a contentious issue as the calculation of the CF is complex and there is ample scope for misleading an uneducated public.

Carbon Labelling (CL) is a system that could enable the consumer to judge the green credentials of a product from the packaging in a similar manner to that for nutritional information for foods. The form of the CL and the method for determining it are again open to interpretation and not as yet subject to international coding or scrutiny although ecolabels are given to products considered to have less environmental impact than their competitors. Such labels take several forms: self-certification by producers, certification by producer trade organisations or by third party (independent) schemes. In all cases the extent of compliance, trade restrictions if products are not certified and the scientific basis for the claims are all contentious issues. The developed economies have driven Ecolabelling as an issue, as one would expect, and companies use it as a marketing tool for their products. One very high profile ecolabelling example is shown in the fish processing industry for canned tuna, although the issues raised have general applicability to the CMI. Thus, Iles (2007) argued that seafood producers are invisible to the consumer who only recognised the well known retailers of their products so consumer pressure on producers was less telling than that of the production chain community (see The Supply Chain below). However, the demand by developed country consumers for ethically sourced fish products has now been extended to include the processing methods used. These should be equally ethical with due respect to the environment and societal impacts employed by the industry. Thrane, Ziegler, and Sonesson (2009) discussed the use of ecolabelling for wild-caught seafood products where LCA studies showed that significant environmental impacts were attributable to post-landing processing and criteria reflecting this should be included in an LCA and energy consumption, materials and waste handling and waste water were highlighted. Although specifically describing a sector of the food processing industry these comments (on producer invisibility and on ethical processing) apply widely in the

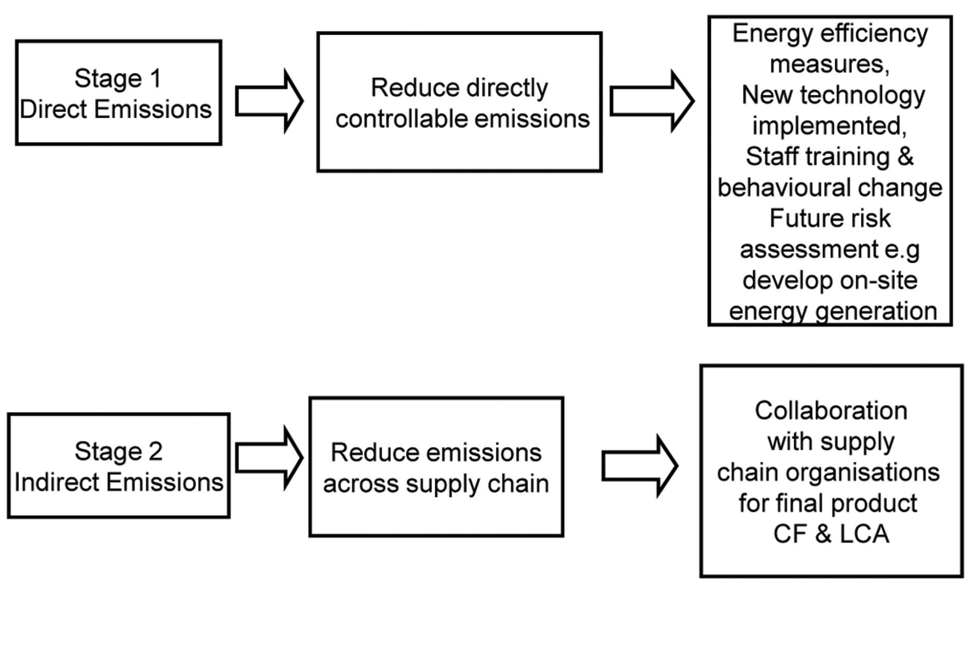
CMI and it would be to the sector's benefit to be less invisible and to promote its ethical processing credentials.

Supply Chain Contribution

Large companies with many business interactions have used interventions in the supply chain to streamline operations and gain commercial benefits through: increased efficiency and productivity; product development and reduced waste. A reasonable question to ask is, can the same approach reduce GHG emissions and promote sustainability? The answer is undoubtedly, "Yes," but when GHG emissions, an LCA or a CF are taken into account current operations along the supply chain might require change to have a positive environmental impact. Actions to reduce the CF, such as energy use reductions, will also give overt economic benefits (or might not be considered for implementation at all) but can also contribute to good public relations (contributing to CSR and the TBL). The supply chain approach must be applied in an all-embracing manner, rather than each company in the chain (including the central operation) looking only at the contribution of their own activities with a cumulative effect—which would be the traditional way to proceed. Such a co-ordinated approach demands collaboration (and trust) up and down the supply chain, around the central operation, with savings being identified for the product as a whole. The Carbon Trust in the UK is one organisation which has developed supply chain models and supported case studies (Carbon Trust, 2006). Figure 4 illustrates the components of the supply chain carbon savings methodology.

Once again the food industry which processes primary products for the consumer is a good example of the supply chain approach. The emissions associated with supplying food to the plate can be divided into direct emissions from energy consumed in the home (23%), indirect emissions from the supply chain (69%), and travel emissions in getting the food to the home (8%) (Carbon Trust, 2006). Thus, the large supply chain component, if mobilised appropriately, could have a massive sustainability impact and the central

Figure 4. Supply chain carbon reduction methodology (a possible stage 3 carbon offsetting has not been included)



company by influencing its suppliers could have global impact and deliver genuine TBL benefits. Iles (2007) argued that seafood producers could be made more accountable through a production chain view and associated pressures making them more transparent in the process and ways to achieve this, were:

- identify and track companies to remove their invisibility
- develop product chain campaigns so that companies influenced each other
- develop mechanisms to compare companies to improve industry practices
- develop methods to track consumption, production and management changes
- develop interactive consumer tools so that consumers get feedback on their purchasing habits which can also be fed back to the producers.

These considerations can be applied across the CMI with equal force.

Reporting Sustainability Activity and CSR

In order to establish the sustainability of the CMI it is necessary to determine the criteria by which it is judged. These criteria might be the same as for any other sector in most cases but should include those peculiar to the CMI. Furthermore it is necessary to apply methodologies which allow measurement of the industry impacts on the environment and the effectiveness of sustainability policies when applied, such as LCA, CF and Supply Chain analysis. In an E-LCA the conventional Impact Analysis categories described previously: global warming; acidification; eutrophication; ozone depletion, land use, and photochemical smog work very well for the CMI as they reflect the importance of emissions and their impact on the environment. However, we have already indicated the need to extend the range of EIA categories to reflect better the impact of the industry through human and eco-toxicology for example. Other categories could reflect the social impact of the CMI, either, directly on workers in the industry, on the surrounding community and even (inter) nationally – the S-LCA. Such categories might be more difficult to define but will allow the LCA approach to reflect all aspects of the TBL rather than just the environmental. One attempt to achieve such reporting is the Global Reporting Initiative (GRI) which has developed sustainability reporting guidelines to encompass the three areas of the TBL—the current guidelines are the third generation (G3) and still subject to scrutiny and change. Starting from a number of common principles the reporting framework (based on the G3 guidelines) includes specific sector supplements to reflect unique sectoral issues and community impacts. It is also developing supply chain sustainability issues through a Global Action Network (GAN) project (Global Reporting, 2011).

A fundamental question raised earlier is whether there can be a coherent approach to sustainability in the CMI sector and, if not, is this important? The development of a standardised set of sustainability criteria would have benefits such as:

- making realistic comparisons between companies and different sub-sectors
- ensuring robust methodologies of assessment through application in a variety of settings
- making knowledge transfer from industry-to-industry or academia-to-industry easier
- establishing a common language of sustainability to be used in negotiations and enforcement of (inter)nationally agreed treaties and protocols
- giving legitimacy to CSR and make communication with the lay public more transparent

- reducing the cost of producing LCA through economies of scale and standardisation of inputs with IT-based support.

At the same time the different scale of operation and range of processes and products within the CMI sector militates against a rigid approach and some sub-sectors face specific sustainability issues which must be addressed in a unique manner, as mentioned above in relation to the GRI sector supplements. For example, while everyone would agree that all products must comply with health and safety criteria in their use the demands on the pharmaceutical industry are greater than on the consumer goods industry which would be greater than on speciality and base chemicals which are more divorced from the consumer. The industry response to these demands must affect the sustainability of that sub-sector and hence the LCA criteria applied to it should reflect this.

Food Processing is one of the specific sectors covered and the main sector topics for reporting developed in a two year process are:

- Sourcing
- Labour/management relations
- Healthy and affordable food
- Public policy
- Costumer health and safety
- Product and service labelling
- Marketing communications
- Breeding and genetics
- Animal husbandry
- Transportation, handling, and slaughter.

The topics listed above reflect the particularly sensitive business of food production and the consumer's perception of the industry where expectations are becoming more and more stringent – but not exclusively so. Sector supplements exist for the energy utilities and Mining and Metals and other CMI sectors with reports in preparation are: construction, oil and gas, the automotive industry and telecommunications.

Green Engineering

Garcia-Serna *et al.* (2007) give a broad review of the philosophies, disciplines and technologies which have been brought together to apply sustainability to chemical (process) engineering and which have been labelled, "Green Engineering." They also recognised that the education of chemical engineers in green engineering was crucial in making the shift to a sustainable future. If green engineering is to be considered at all levels (molecular, micro- and macro-) it should be included from the early stages of the educational programme. Graedel and Allenby (2009) suggested that green engineering is a move towards a more responsible technology but did not ask how the social and environmental aspects of sustainability can be incorporated into the engineering profession.

The American Chemistry Society (ACS) delivered a public policy statement on sustainability and the, "chemical enterprise," which consists of all the industry, trade associations and educational and professional organisations underpinning the sector suggesting that they had a role to play in sustainable development (American Chemistry Society, 2008). The document identified eight areas of importance for a sustainable chemical enterprise:

- green sustainable chemistry
- LCA
- toxicology
- renewable feedstocks
- renewable fuels
- energy intensification of processing
- separation, sequestration and use of carbon dioxide
- sustainability education.

In addition, the non-technical barriers to a sustainable CMI were recognised (Satterfield, et al., 2009):

- develop working definitions and practical metrics to measure progress towards sustainability
- quantify the true cost of products to promote sustainable options
- promote cross-functional and multi-disciplinary communication
- support continuous improvement through forward thinking, collaborative, goal-orientated, non-technology specific regulations and/or incentives which could be adapted as sustainable technology evolves
- incorporate sustainability principles at all levels of education.

The report then went on to recommend various support tools to achieve these aims. Later work by the Harvard-Yale-ACS-GCI Green chemistry project (Matus, et al., 2007) delved deeply into the challenges facing green chemistry and recognised the following barriers to implementation: economic and financial, regulatory, technical, organisational, cultural and definitions and metrics.

Each of these barriers covered a range of objections; for example, included under the regulatory barrier was the possible cost of recertification of processes by the US FDA (Food and Drug Administration) when changes were made to already accepted processes in the name of sustainability. Under cultural barriers were a lack of awareness in both the chemistry community and the general public and common misconceptions that green chemistry products were more expensive, less effective and not based on rigorous science. Some possible solutions to these barriers were mooted, such as: creating incentives for development/implementation of innovation; facilitating linkages and networks to spread the word on green chemistry; increase research activity in green chemistry and to raise the profile of positive environmental/health impacts rather than taking a defensive approach. Finally, there needed to be a national framework for green chemistry policy which should reach out beyond the USA, particularly to the emerging chemical enterprise in China and India and it falls to the process engineering fraternity to do likewise and deliver a CMI that reflects these changes.

Public Perception and Ethical Considerations

The public perception of the CMI as described in the Introduction to this chapter is usually one of a disaster waiting to happen. Pollution, spillages and explosions when they occur are headline news yet the safe operation of processing plant, governed by strict Health and Safety procedures, goes unmentioned and unappreciated. This might seen to be the accepted view of the news industry, after all, "Man bites dog" will always outweigh, "Dog bites man." As mentioned previously proactive reporting of sustainable achievements and CSR are the way forward for the sector rather than the current invisibility—the public needs to acknowledge the debt it owes to the CMI for making life more than tolerable for many people. While it will always be a difficult sell—few processing plants are easy on the eye, nose or ear—an open, honest approach to immediate neighbours and the wider community is essential. It must also be recognised that the current fossil fuel-based CMI did not spring up fully formed but has evolved over the last 150 years and with each step change in processing plant complexity and process innovation mistakes, sometimes catastrophic, have been made. Modern computer-aided design and modelling have taken some risk out of plant and process development but the keen eye of the experienced process engineer is still needed to spot potential design faults.

The future of the CMI will throw up new challenges of both public perception and ethics when the sector becomes reliant on renewable biomass as a raw material and on synthetic biology as a means of its conversion to useful products. The control of renewable resources may benefit some countries through their ownership of vast biomass reserves (e.g. Brazil) but some smaller countries will find it hard to protect their natural resources (e.g. policing activity at sea even within a 200 mile Exclusive Economic Zone). The prospect of synthetic biologists, "playing God," by creating new life forms and the possible dangers of accidental or criminal release of such life forms is already exercising the minds of legislators and the scientific community alike. Given the past experience of the CMI as a fossil fuel-based enterprise it seems incumbent upon it to tackle the new challenges openly and to equip the process engineer of tomorrow to understand the issues. Some comments on these issues will be made in the sections below on the new technologies for the CMI.

New Technology Affecting the CMI

The development of a sustainable CMI demands the application of new technologies and fortunately process engineering has responded in the past to such demands (Favre, Falk, Roizard, & Schaer, 2008) and it is suggested that they have an ethical and moral duty to do so again to deliver a sustainable society (Byrne & Fitzpatrick, 2009). What are the challenges for a sustainable CMI? Firstly, the need to recognise that the CMI cannot rely on the traditional fossil sources of basic chemicals for manipulation and transformation into useful products and that the process operations based on these raw materials are not sustainable. Instead the CMI will rely on renewable raw materials based on biomass for platform chemicals, transforming them into products (currently focussed on polymers and fuels) which are described as, "bio-based," materials. This will have fundamental implications for sustainable engineering based on renewable carbon accounting and on claims to be a carbon-neutral industry. Additionally, it will be necessary to take the present state of knowledge in biotechnology to another level where engineering and biology are combined to design and synthesise novel functions and systems ranging from enzyme combinations to whole organism analogues—synthetic biology. The ability to design and synthesise DNA sequences and synthetic biological systems based on them will bring a completely new view of process engineering with governance, intellectual property, safety/security, and ethical issues demanding a new type of process engineer to deal with them including the ability to grasp the wider social dimensions of developing processes/products as proposed under the S-LCA banner. It has been suggested that these pressures could lead to the next paradigm shift for process engineering (Hall & Howe, 2010). The implications of a bio-based CMI leaning heavily on synthetic biology to replace fossil fuel technology are profound and worthy of more detail here.

Development of a Bio-Based CMI

The biggest current use of biomass is for conversion to biofuels although this is not the most efficient use for it and does not reflect the large number of fossil fuel-based products which must be substituted for in the future. So, biomass could be the source of plastics, paints and dyes, adhesives and fertilisers to name a few examples. The biomass available for conversion is made up of cellulose, hemicellulose and lignin which are usually found in conjunction in plant material. Cellulose is the easiest molecule to manipulate, being a polymer of glucose, whilst hemicellulose and lignin are more intractable materials and make cellulose harder to extract. Sources of biomass include natural forests, grasslands and marine ecosystems and agro-systems such as field crops and plantations. The commercial exploitation of these sources; their geographical location and ownership and their relationship to human food production are all fiercely debated issues where the ethics and CSR for the CMI are under heavy pressure. The term biomass also gives the idea that all is grist to the mill and the biodiversity of the parent sources is overlooked although they could be the source of valuable natural products, livelihoods and cultural significance for the local population (Erosion, Technology and Concentration, 2010).

Synthetic Biology

Relationships between engineering and biology have developed over a long time: from the early food fermentations through to modern industrial biotechnology using whole organisms or enzyme extracts to produce products such as antibiotics and functional food components. These activities have also brought about a good understanding of the physiology, growth characteristics and genetics of the organisms involved, such as *Penicillium* spp, *Aspergillus* spp, *Saccharomyces cerevisiae,* and *Escherichia coli* which could be manipulated for greater expression of natural or implanted products. The genetic manipulation of these organisms in current biotechnology is considered to be *ad hoc* or serendipitous at best whereas synthetic biology offers specific design for purpose.

Although it has grown from the engineering/biology axis other fundamental advances in a third sphere have enabled synthetic biology to become nearer to reality. These developments in mathematics, computer modelling and information storage, retrieval and communication and have enabled the complex biological systems to be mapped, quantified and masses of data transmitted as necessary. As a result of the potential control which can be achieved there are great claims for synthetic biology in fulfilling the promises which conventional biotechnology cannot deliver. Table 2 gives a list of the broad areas of activity with specific applications within these areas. The list is pretty all-encompassing with no area of current process engineering omitted. Immediate goals are the next generation of biofuels which are capable of driving the high-fuel usage sectors of heavy haulage, railways and aviation. Another area of immediate impact will be in the production of drugs and pharmaceuticals such as the production of the antimalarial drug (artemisinin) precursor, artemisinic acid in engineered yeast (Ro, et al., 2006) as an alternative route to extraction from the medicinal plant, where limited availability is reflected in price, and a shift from plant to microbial production would give a reliable and affordable supply (Zeng, Qiu, & Yuan, 2008). This project was supported by the Bill and Melinda Gates Foundation, addresses a massive health issue and being near to fruition thus attracts the headlines for synthetic biology.

Synthetic biology is also of great public interest because it deals in the fundamentals of life with concomitant ethical, security and safety issues. The successful sequencing of the human genome (International Human Genome Sequencing Consortium, 2001; Venter, et al., 2001) gave synthetic biology another media boost but also raised the spectre of, “Frankenstein,” organisms with malign uses or, at the very least, the threat of accidental release. Thus, this eye-catching and newsworthy disciple has been the subject of debate far removed from academia (Specter, 2009) which has brought about calls for regulation by bodies other than the active synthetic biology commu-

Table 2. Applications for synthetic biology

Energy	Bio-medical	Environment	Agriculture	Other
Biofuels	Smart drugs & delivery systems	New (biodegradable) packaging	Biomass utilization	Smart materials
Cellulosic biomass conversion	Therapy vectors	Spillage detection	Food production	Green chemistry & engineering (sustainability)
	Tissue repair	Bioremediation		Synthesis of non-natural molecules (e.g. proteins)

nity itself. Alongside the scientific discovery goes great commercial potential and fears for a profit-driven monopoly of multinational companies who can afford to fund the initial research (Erosion, Technology and Concentration, 2010). The replacement of natural sources of bio-active compounds (such as artemisinin) could also deprive developing countries of revenue (Balmer & Martin, 2008). With synthetic biology being based on the manipulation of genetic material there is the added issue of the Intellectual Property Rights (IPR) for the new life forms created. The models for the development of the synthetic biology-based industry may learn lessons from the fast-changing computing and software industry (Henkel & Maurer, 2007).

The pace of progress in synthetic biology is also frightening with the first synthetic organism, from the poliovirus genome (7,741 base pairs) taking two years of work (Cello, Paul, & Wimmer, 2002) whilst the genome of the bacteriophage ΦX-174 (5,386 base pairs) took just two weeks in the J Craig Venter Institute (Smith, Hutchinson, Pfannkoch, & Venter, 2003). The first bacterium to be synthesised (also by the Venter Institute) was *Mycobacterium genitalium* with 589,000 base pairs (Gibson, et al., 2008). This opened up the concept of a viable minimal cell, produced by removing genes which are not essential for life, but with the minimum number of components to support biological synthesis from inserted DNA material. *M genitalium* carries 485 genes and has the smallest genome in laboratory organisms and about one quarter (115 or more) could be removed with no affect on genome functionality. More conventional microbial carriers (or "chassis") for synthetic biology include E coli, yeasts, *Bacillus subtilis,* and *Pseudomonas putida* reflecting the wide knowledge about these organisms built up over the years. Future developments could include cell-free applications where the non-specific elements of the living cell chassis are removed from the application.

Synthetic biology owes a debt to the engineering design cycle approach for its development. Process engineers are well versed in taking concepts from the chemistry and biochemistry fields and converting them into industrial scale processes through a process of design-build-commission. Plant design involves putting together a multitude of standard parts and unit operations into the most efficient system through an iterative process—recognising that a number of designs and process routes might be feasible practically but not from a commercial point of view. Learning the art of process design is at the very core of chemical engineering education being the culmination of the separate disciplines which make up the subject. Thus, process design is based on fundamental concepts such as heat and mass transfer, transport phenomena, reaction kinetics and modelling and fluid mechanics. The basic principles for the process in question are then broken down into the unit operations which comprise the complete process and described in block flow diagrams, process flow diagrams and piping and instrumentation diagrams. Other considerations include product throughput and quality; financial constraints and process criteria such as safety, flexibility and operational characteristics. All this knowledge can be applied to synthetic biology in the consideration of what route to take to production. The analogy between (process) engineering design and synthetic biology has been writ large in that the idea of taking discrete and standardised genetic units and putting them together to make an engineered organism is as direct or simple as in the physical world. One approach to the provision of these standard parts is the Biobricks Foundation which is a public-benefit organisation originating from workers at MIT, Harvard and University of California, San Francisco. The declared aim of the foundation is to make freely available the fundamental components, or bioparts, needed to conduct synthetic biology and in doing so to introduce the mindset which underpins, for example, chemical engineering design (Biobricks Foundation, 2011).

However, just as physical chemical plant must be coaxed into an operational equilibrium so the genomic components of a synthesised organism must be brought together carefully to function correctly and there has been an evolution in the complexity and scale at which this can be done (Andrianantoandro, Basu, Karig, & Weiss, 2006; Canton, Labno, & Endy, 2008; Purnick & Weiss, 2009).

The other, more practical, input from the chemical engineering fraternity to synthetic biology could be in defining the processability of engineered organisms and their products. Currently synthetic biology seems to be directed towards the genetic control of the organism for specific product expression but a substantial element in the cost of bio-products is their separation and purification—Downstream Processing (DSP). Here the chemical engineer could ask the question about processability and define the DSP characteristics required in the final design. This issue is being addressed in some instances, for example the use of *e coli* to produce biofuels which are expressed externally and being immiscible with the aqueous fermentation broth can separated from it easily. Where products are expressed and water soluble it may be possible to select for ease of DSP, for example, by developing organisms which require a minimal/simple fermentation medium which can be easily removed in the early/high volume stages of DSP. The minimal cell would be a good model for this approach. Where products are expressed as internal inclusion bodies they must be separated from the cell contents. Ways to achieve this separation would be helped by simple cell biochemistry and susceptibility to cell disruption paving the way for separation of the cell contents and the inclusion body.

Education for Sustainable Process Engineering

There is a need to ensure that the new generation of chemical engineers recognises the importance of sustainability in general and how to apply such principles in all aspects of process design. Garcia-Serna *et al.* (2007) proposed that sustainability should be introduced early on and so permeate all the other major teaching elements which make up a chemical engineering course - but how should sustainability be incorporated into a very full curriculum and at what level (see Table 3)?

Other aspects of a university degree have always been prominent in chemical engineering education such as problem solving and the modern chemical engineer with good mathematical ability and literacy, combined with the ability to communicate with those from a range of disciplines, can find employment in areas far removed from the CMI. The issue of how to fit sustainability into this crowded timetable offers a range of options (Favre, et al., 2008):

Table 3. Chemical engineering degree curriculum

Curriculum area	Subject
Basic Science	Mathematics
	Physics
	Chemistry
	Computing
Core engineering	Thermodynamics
	Physical chemistry
	Fluid mechanics
	Transport phenomena
	Unit operations
	Reaction engineering
	Plant design
	Process dynamics & control
	Health, safety & environment
	Plant, equipment & materials
Electives	Biotechnology, bioprocessing & fermentation
	Particle science
	Polymer technology

- no change in the curriculum as it fits the bill already and has proved to do so despite significant changes in the industry
- overhaul the curriculum to become a, "curriculum for the future" (Armstrong, 2006)
- adapt the current curriculum to meet the new challenges of sustainability and the needs of green chemistry (McDonough, Braungart, Anastas, & Zimmerman, 2003).

A recent exercise to establish how sustainability, in the broadest sense, is being incorporated in to chemical engineering education was conducted in the USA (Allen, Murphy, Allenby, & Davidson, 2009a) covering 366 engineering colleges and all the engineering disciplines including civil, environmental, mechanical and systems engineering and also chemical, bio and materials engineering. Four strategies were discerned:

- dedicated sustainable engineering courses in any of the major disciplines
- integrate sustainable engineering concepts into existing courses to raise awareness of students
- focus on technologies which will be important in sustainable engineering
- mixed teaching strategy and some interdisciplinary courses with non-engineering departments

Allenby, Murphy, Allen, and Davidson (2009b) reflected on the conceptual challenge of sustainable engineering and the educational conflict between teaching the topic as a speciality or as a component of other engineering disciplines. They commented on the social dimensions of sustainability which are less quantifiable and more normative and subjective than the typical material taught in engineering courses and which staff found less easy to teach.

Dedicated undergraduate sustainable engineering courses might only attract those students for whom the very idea of sustainability is vitally important whilst deterring the general engineer. A standard engineering degree in any subject area, including sustainability elements, followed by specialisation at post-graduate level might be a better way to attract engineers into the area of sustainability. Another argument says that all engineers should be aware of sustainability, hence the need for integrated courses but a limited number need to be specialists – hence the need for post-graduate courses in sustainability (with proper focus). As mentioned above the current chemical engineering degree course is full of intellectually-demanding material which is taught through a problem solving/design format combined with practical classes and students finding their way through this challenge become well-rounded individuals capable of tackling the wide sustainability agenda at this time. In this case the mixed teaching strategy would come to the fore but this leads the education debate into the realms of the disciplinarity of sustainable develop.

Chemical Engineering in the Future: Disciplinarity

The by-word for sustainability has been cross- or multidisciplinarity as demonstrated by the TBL or concentric models (Satterfield, et al., 2009). However, within academia the, "silo," mentality still prevails with fierce protection of expertise or specialist knowledge and a resultant lack of understanding between the social, cultural, scientific and economic disciplines and most importantly what can be learned by one discipline from the others. There is also a fear factor in entering into another domain where the academic is not the master and their assumptions and knowledge can be questioned. But technical solutions to sustainability can only work if they have social and cultural acceptance and their economic significance is recognised—the TBL and S-LCA writ large. Academics who move from their own domain into that of others can be deeply mistrusted and treated as dilettantes at best and loose cannons at worst. Yet it is such free thinkers who can draw together the various strands of the sustainability web and recognise the interconnectivity of the subject. Another issue working against the cross- or multidisciplinary approach to sustainability is early specialisation which often separates the arts and sciences in the secondary schooling system and leads the science student into ever more esoteric areas as they progress through the university system. This experience does not develop the wide-ranging broad-brush mental flexibility capable of dealing with sustainability (and with synthetic biology) and as a result regards it as a "fuzzy" or "elusive" concept, in particular its qualitative aspects. The gulf between the "two cultures" of the social and physical sciences described by C. P. Snow (1959) still holds today.

CONCLUSION

In the future, the concepts of sustainability must be embraced and their application in the CMI must be articulated to a sceptical public—the sector must be seen to be doing its bit. The advent of a new raw material base, and the means to manipulate it, must be welcomed but will mean an equally new paradigm to process engineering practice. The meeting of biology, engineering and information sciences (in synthetic biology) will demand a flexible approach, one which the process engineering curriculum and design teaching already supports. The commercial, ethical, and social impact of a sustainable CMI brought about by this new world will be under increasing scrutiny and an open, honest relationship with the public should be encouraged, indeed celebrated.

REFERENCES

Allen, D. T., Murphy, C. F., Allenby, B. R., & Davidson, C. I. (2009a). Incorporating sustainability into chemical engineering education. *Chemical Engineering Progress*, *105*(1).

Allenby, B. R., Murphy, C. F., Allen, D. T., & Davidson, C. I. (2009b). Sustainable engineering education in the United States. *Sustainability Science*, *4*, 7–15. doi:10.1007/s11625-009-0065-5

American Chemistry Society. (2008). *Statement on sustainability of the chemical enterprise*. Washington, DC: ACS.

Andersson, K., Ohlsson, T., & Olsson, P. (1998). Screening life cycle assessment (LCA) of tomato ketchup: A case study. *Journal of Cleaner Production*, *6*, 277–288. doi:10.1016/S0959-6526(98)00027-4

Andrianantoandro, E., Basu, S., Karig, D. K., & Weiss, R. (2006). Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology*, *2*, 1–14. doi:10.1038/msb4100073

Aragon-Correa, J. A., & Sharma, S. (2003). A contingent resource-based view of proactive corporate environmental strategy. *Academy of Management Review*, *28*(1), 71–88.

Armstrong, R. C. (2006). A vision of the curriculum of the future. *Chemical Engineering Education: Official Journal of the European Federation of Chemical Engineers*, *D*(1), 104–109.

Azapagic, A., Clift, R., & Predan, S. (2004). *Sustainable development in practice - Case studies for engineers and scientists*. New York, NY: John Wiley and Sons. doi:10.1002/0470014202

Balmer, A., & Martin, P. (2008). *Synthetic biology: Social and ethical challenges*. Swindon, UK: BBSRC.

Batterham, R. J. (2006). Sustainability – The next chapter. *Chemical Engineering Science*, *61*(13), 4188–4193. doi:10.1016/j.ces.2005.10.016

Benoit, C., & Mazijn, B. (Eds.). (2009). *Guidelines for social life cycle assessment of products*. Geneva, Switzerland: United Nations Environment Programme.

Benoit, C., Norris, G. A., Valdivia, S., Ciroth, A., Moberg, A., & Bos, U. (2010). The guidelines for social life cycle assessment of products: Just in time. *International Journal of Life Cycle Assessment*, *15*, 156–163. doi:10.1007/s11367-009-0147-8

Biobricks Foundation. (2011) *Website*. Retrieved from http://biobricks.org/.

Boustead, I. (1996). LCA - How it came about: The beginning in the UK. *International Journal of Life Cycle Assesment*, *1*, 147–150. doi:10.1007/BF02978943

Brown, D., Dillard, J., & Scott Marshall, R. (2006). *Triple bottom line: A business metaphor for a social construct*. Barcelona, Spain: Universitat Autònoma de Barcelona.

Byrne, E. P., & Fitzpatrick, J. J. (2009). Chemical engineering in an unsustainable world: Obligations and opportunities. *Education for Chemical Engineers*, *4*, 51–67. doi:10.1016/j.ece.2009.09.001

Canton, B., Labno, A., & Endy, D. (2008). Refinement and standardisation of synthetic biological parts and devices. *Nature Biotechnology*, *26*, 787–793. doi:10.1038/nbt1413

Carbon Trust. (2006). *Carbon footprints: In the supply chain, the next step for business*. London, UK: Carbon Trust.

Carroll, A. B. (1999). Corporate social responsibility: Evolution of a definitional construct. *Business & Society*, *38*, 268–295. doi:10.1177/000765039903800303

Cello, J., Paul, A. V., & Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science*, *297*(5583), 1016–1018. doi:10.1126/science.1072266

Choi, B.-C., Shin, H.-K., Lee, S.-Y., & Hur, T. (2006). Life cycle assessment of a personal computer and its effective recycling rate. *International Journal of Life Cycle Assessment*, *11*(2), 122–128. doi:10.1065/lca2004.12.196

Coley, N. G., & Wilmot, S. A. H. (2000). Chemical industry and the quality of life In Russell, C. A. (Ed.), *Chemistry Society and Environment: A New History of the British Chemical Industry* (pp. 319–349). Cambridge, UK: Royal Society of Chemistry.

Erosion, Technology and Concentration. (2010). *The new biomassters, synthetic biology and the next assault on biodiversity and livelihoods*. ETC Group Communiqué #104. Ottawa, Canada: ETC.

Evans, G. M., Galvin, K. P., & Doroodchi, E. (2008). Introducing quantitative life cycle analysis into the chemical engineering curriculum. *Education for Chemical Engineers*, *3*, 57–65. doi:10.1016/j.ece.2008.01.003

Favre, E., Falk, V., Roizard, C., & Schaer, E. (2008). Trends in chemical engineering education: Process, product and sustainable chemical engineering challenges. *Education for Chemical Engineers*, *3*, 22–27. doi:10.1016/j.ece.2007.12.002

Florida, R. (1996). Lean and green: The move to environmentally conscious manufacturing. *California Management Review*, *39*(1), 80–105.

Garcia-Serna, J., Perez-Barrigon, L., & Cocero, M. J. (2007). New trends for design towards sustainability in chemical engineering: Green engineering. *Chemical Engineering Journal*, *133*, 7–30. doi:10.1016/j.cej.2007.02.028

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., & Zaveri, J. (2008). Complete chemical synthesis, assembley and cloning of a mycoplasma genitalium genome. *Science*, *319*(5867), 1215–1220. doi:10.1126/science.1151721

Global Reporting. (2011). *Global action network for transparency in the supply chain*. Retrieved from http://www.globalreporting.org/CurrentPriorities/supplyChain/GlobalActionNetwork/.

Graedel, T. E., & Allenby, B. R. (2009). *Industrial ecology and sustainable engineering*. New York, NY: Prentice Hall.

Griesshammer, R., Benoît, C., Dreyer, L., Flysjo, A., Manhart, A., Mazijn, B., et al. (2006). *Feasibility Study: Integration of social aspects into LCA*. Retrieved from http://www.concisenet.de/fileadmin/download/modul_ps/UNEP-SETAC_feasibility_study_mai_06.pdf.

Hall, G. M., & Howe, J. M. (2010). Sustainability of the chemical manufacturing industry – Towards a new paradigm? *Education for Chemical Engineers*, *5*, 100–107. doi:10.1016/j.ece.2010.09.001

Henkel, J., & Maurer, S. M. (2007). The economics of synthetic biology. *Molecular Systems Biology*, *3*(117), 1–4.

Iles, A. (2007). Making the seafood industry more sustainable: Creating production chain transparency and accountability. *Journal of Cleaner Production*, *15*, 577–589. doi:10.1016/j.jclepro.2006.06.001

International Standards Organisation. (2006). *International standards organisation, environmental management – Life cycle assessment – Principles and framework*. ISO 14040:2006 (E) Series. Geneva, Switzerland: ISO.

Jorgensen, A., Finkbeiner, M., Jorgensen, M. S., & Hauschild, M. Z. (2010). Defining the baseline in social life cycle assessment. *International Journal of Life Cycle Assessment*, *15*, 376–384. doi:10.1007/s11367-010-0176-3

Judge, W. Q., & Douglas, T. J. (1998). Performance implications of incorporating natural environmental issues into the strategic planning process: An empirical assessment. *Journal of Management Studies*, *35*(2), 241–260. doi:10.1111/1467-6486.00092

Jungmeier, G., Werner, F., Jarnehammar, A., Hohenthal, C., & Richter, K. (2002). Allocation in LCA of wood-based products: Part I methodology. *International Journal of Life Cycle Assessment*, *7*(5), 290–294. doi:10.1007/BF02978890

Labuschagne, C., Brent, A. C., & Van Ercka, R. P. G. (2005). Assessing the sustainability performance of industries. *Journal of Cleaner Production*, *13*, 373–385. doi:10.1016/j.jclepro.2003.10.007

Matus, K. J. M., Anastas, P. T., Clark, W. C., & Itameri-Kinter, K. (2007). *Overcoming the challenges to the implementation of green chemistry*. CID Working Paper No. 155. Cambridge, MA: Harvard University.

McGee, J. (1998). Commentary on corporate strategies and environmental regulations: An organizing framework. *Strategic Management Journal*, *19*, 377–387. doi:10.1002/(SICI)1097-0266(199804)19:4<377::AID-SMJ988>3.0.CO;2-S

Ntiamoah, A., & Afrane, G. (2008). Environmental impacts of cocoa production and processing in Ghana: Life cycle assessment approach. *Journal of Cleaner Production*, *16*, 1735–1740. doi:10.1016/j.jclepro.2007.11.004

Purnick, P. E. M., & Weiss, R. (2009). The second wave of synthetic biology: From modules to systems. *Nature Reviews. Molecular Cell Biology*, *10*, 410–422. doi:10.1038/nrm2698

Ro, D., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., & Ndungu, J. M. (2006). Production of the antimalarial drug precursor atemisinic acid in engineered yeast. *Nature*, *440*, 940–943. doi:10.1038/nature04640

Satterfield, M. B., Kolb, C. E., Peoples, R., Adams, G. L., Schuster, D. S., & Ramsey, H. C. (2009). Overcoming nontechnical barriers to the implementation of sustainable solutions in industry. *Environmental Science & Technology*, *42*, 4221–4226. doi:10.1021/es802980j

Skals, P. B., Krabek, A., Nielsen, P. H., & Wenzel, H. (2008). Environmental assessment of enzyme assisted processing in pulp and paper industry. *International Journal of Life Cycle Assessment*, *13*(2), 124–132. doi:10.1065/lca2007.11.366

Smith, H. O., Hutchinson, C. A., III, Pfannkoch, C., & Venter, J. C. (2003). Generating a synthetic genome by whole genome assembly: ΦX174 bacteriophage from synthetic oligonucleotides. *National Academy of Sciences, 100*(26), 15440-15445.

Snow, C. P. (1959). *The two cultures*. Cambridge, UK: Academic Press.

Specter, M. (2009, September 28). A life of its own: Where will synthetic biology lead us? *The New Yorker*.

Thrane, M., Ziegler, F., & Sonesson, U. (2009). Eco-labelling of wild-caught seafood products. *Journal of Cleaner Production*, *17*, 416–423. doi:10.1016/j.jclepro.2008.08.007

United Nations Environment Programme. (1996). *Life cycle assessment: What it is and how to do it*. New York, NY: United Nations Environment Programme (UNEP).

Venter, J. C. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. doi:10.1126/science.1058040

Weidema, B. (2003). *Market information in life cycle assessment*. Environmental Project No 863 2003. Copenhagen, Denmark: The Danish Environmental Protection Agency.

Wernet, G., Conradt, S., Isenring, H. P., Jimenez-Gonzalez, C., & Hungerbuhler, K. (2010). Life cycle assessment of fine chemicals production: A case study of pharmaceutical synthesis. *International Journal of Life Cycle Assessment*, *15*(3), 294–303. doi:10.1007/s11367-010-0151-z

Wiedmann, T., & Minx, J. (2007). *A definition of carbon footprint*. ISA^{UK} Research Report 07-01. London, UK: Integrated Sustainability Analysis.

World Commission on Environment and Development. (1987). *Our common future*. Oxford, UK: Oxford University Press.

This work was previously published in Global Sustainable Development and Renewable Energy Systems, edited by Phillip Olla, pp. 131-151, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 89
Cellular or Functional Layout?

Abdessalem Jerbi
University of Sfax, Tunisia

Hédi Chtourou
University of Sfax, Tunisia

ABSTRACT

The cellular layout has been compared to the traditional functional layout using multiple comparison methodologies that either lack objectivity or are highly time-consuming. The main purpose of this chapter is to propose a novel and objective methodology. Hence, a critical analysis of ten comparison studies is followed by the presentation of the layouts simulation models. Subsequently, the proposed comparison methodology is described. Following this methodology, simulations are conducted according to a plan of experiments developed from Taguchi standard orthogonal arrays. Consequently, results, expressed in Signal to Noise ratios, are analyzed using ANOVA. Next, a mathematical model is derived by interpolation between the factors and interactions effects. This model must be validated by the confirmation test, otherwise the comparison methodology should be reiterated while considering new interactions. This cycle should be reiterated as much as necessary to obtain a valid mathematical model. The proposed comparison methodology has been applied with success on an academic manufacturing system.

INTRODUCTION

The increased competition within industry has resulted in manufacturing companies spending considerable effort to improve flexibility and responsiveness to meet customer needs. Cellular manufacturing, a facet of group technology, has emerged as one of the major techniques being used for the improvement of manufacturing competitiveness. A large number of empirical, analytical and simulation studies have been devoted to compare the cellular layout (CL) to the classical functional layout (FL). Simulation-based comparative studies constitute the mainstream of this research field. Varied results were reported by these comparative simulation studies. Indeed, different researches found the FL always superior to the CL with regard to all used performance measures (Jensen, Malhotra, & Philipoom, 1996; Morris & Tersine, 1990, 1994). Further researches reported that the CL is superior to the FL in all operating conditions (Pitchuka, Adil, & Anantha-

DOI: 10.4018/978-1-4666-1945-6.ch089

kumar, 2006; Shafer & Charnes, 1992). Finally, other simulation studies showed that every layout could outperform the other in different particular experimental conditions (Faizul huq, Douglas, & Zubair, 2001; Farrington & Nazametz, 1998; Li, 2003; Shafer & Charnes, 1995; Suresh & Meredith, 1994). The divergence in the studies conclusions is referred to as the “cellular manufacturing paradox” (Shambu, Suresh, & Pegels, 1996). In fact, Agarwal and Sarkis (1998) and Shambu et al. (1996) reviewed a number of FL-CL comparative studies. However they did not identify any objectivity flaws responsible for the conflicting conclusions. Indeed, they simply reported the major findings of some published studies without any critical objectivity assessment.

Actually, methodologies used by comparison studies vary widely but can be classified into three groups. In the first group, authors used the one-factor-at-a-time method. So the two layouts are first compared for one manufacturing context considered as the “base model”. Then, other experiments are carried out in order to test the robustness of the layout choice obtained in the base model. Every experiment corresponds to the modification of a single operating factor (Morris & Tersine, 1990, 1994). In the second group authors considered only some specific combinations of the studied factors settings without any justification (Faizul huq et al., 2001; Li, 2003; Suresh & Meredith, 1994). In the third group authors used the full factorial design technique in order to study the effect of all factors (Farrington & Nazemetz, 1998; Jensen et al., 1996; Pitchuka et al., 2006; Shafer & Charnes 1992, 1995). Methodologies belonging to the two first groups undoubtedly lack objectivity in the choice of the experimentation conditions. Therefore, they do not permit to attach any statistical confidence level to their conclusions. In addition, they do not provide any information about factor interaction. The third group methodology is highly time-consuming. In addition, it is impractical when the number of factors to study is large.

This chapter essentially focuses on the development of an objective FL-CL comparison. It first highlights the lacks of objectivity of the main published FL-CL simulation-based comparison studies in order to explain the origin of their conflicting conclusions. Then it deals with the development of comprehensive FL and CL simulation models using the widely used commercial simulation software Arena 7.0. Finally, it presents the framework of a methodology, based on the coupling of the Taguchi method of experiment design (TM) and simulation. This methodology can be easily applied to any manufacturing context and provides trustworthy results with a minimum experimentation effort.

The remainder of this chapter is organized as follows. The next section presents a taxonomy of the key factors used in the main published FL-CL comparison simulation studies. The foremost used performance measures are also presented in this section. Finally it presents and analyses the findings of a number of relevant studies. The third section presents some general simulation features, needed for modeling both layouts. Then, it respectively gives details of the developed FL and the CL simulation models. Section four gives a general presentation of the objective comparison methodology and then presents a comprehensive academic case study depicting its application. The final section includes some general conclusions and discusses future work prospects.

COMPARATIVE STUDIES FRAMEWORK

Main Experimental Factors

General Manufacturing System (MS) Characteristics

Every MS is characterized by a number of machines arranged either into departments in the functional layout, or else, into manufacturing cells in the cellular layout. Following the FL structure, the shop is composed of *d* departments *Di* (*i*=1, …, *d*). Each of them includes *Mn* functionally equivalent machines. In contrast, the CL is composed of *c* independent manufacturing cells *Cj* (*j*=1, …, *c*). Each cell is a cluster of *Mf* different machines dedicated to a number of similar part types. Furthermore, every MS is designed for a demand pattern comprising different products. Products are identified by two indicators, which are the type (*t*) and the family (*f*). Products are grouped into families according to the similarity of their manufacturing process. Each product type requires a number of manufacturing operations (*mopt*).

Degree of Decomposability of the Part Machine Matrix (DD)

This degree translates the feasibility of the decomposition of the MS into independent cells. In fact, the more the product/machine matrix is diagonal, the more the decomposability is feasible. This degree is negatively correlated to the density of off-diagonal elements.

Batch Size (BS)

Products are generally manufactured and transferred in batches in order to reduce machine setup and transport between machines. Numerous authors included *BS* in there comparison studies as a variable factor and demonstrated that the combination of small batch sizes with an efficient scheduling rule results in the improvement of the cellular layout performances. Most authors used the same batch size for both cellular and functional layouts.

Demand Rate (DEMAND)

The demand rate is mainly expressed by the batch inter-arrival times (*IAT*) in the MS. A large part of authors generated this time by common probabilistic distributions. Others used constant *IAT*. Besides, some authors focus only on the stability of this factor without changing its average value.

Transfer Time (TT)

This parameter corresponds to the interdepartmental travel times in the FL. They are often modeled using appropriate probabilistic laws. In the CL these times correspond to the durations of intra-cell moves. Generally, they are very small compared to those in the FL.

Transfer Mode (TM)

Because of the considerable interdepartmental distances in the FL products are generally transferred by batches in order to reduce transfer costs. Some studies also used this transfer mode between same-cell machines whereas others make use of operations overlapping. This mode exploits the proximity of same- cell machines to allow simultaneous execution of different operations on parts of the same batch.

Flow Direction (FLOW)

A number of authors included the flow direction within a cell as an experimental factor. This factor has two possible levels: “unidirectional” or “backtracking allowed”.

Scheduling Rules (RULE)

Part batches arriving at a department or a cell are put in a waiting queue until the required machine becomes idle. These batches are then sequenced in order to establish the order in which they will be processed. This order is specified by the use of standard scheduling rule such as “First Come First Served” (FCFS), “Shortest Process Time” (SPT), “Earliest Due Date” (EDD) or else, “Repetitive Lots” (RL). The limited versions of the first three rules, FCFS-L, SPT-L and EDD-L are used in order to avoid the duplication of machines setups for the same product type. Finally, the RL rule selects batches of the same type that the one just processed in order to minimize setups.

Processing Time (PT) and Set up Time (ST)

As for the *IAT*, most studies generally modeled both times by independent probabilistic laws. On the other hand, other studies formulated *ST* as a fraction of *PT*.

Set up Time Reduction Factor (δ)

This factor materializes one of the most key advantages of the CL. Indeed, part types of a same family need generally similar setups. Hence, if a machine is set up for a part type and then should be set for a same-family part type, the nominal setup time for the second part is reduced by the δ factor.

Performance Measures

Work in Process (WIP)

WIP is one of the most popular performance measures used in the FL-CL comparative studies (Farrington & Nazametz, 1998; Jensen et al., 1996; Li, 2003; Morris & Tersine, 1990, 1994; Shafer & Charnes, 1992, 1995; Suresh & Meredith, 1994). It essentially characterizes the fluidity of the material flow in the system.

Mean Flow Time (MFT)

MFT constitutes the other most popular measure used in FL-CL comparative studies (Faizul huq et al., 2001; Farrington & Nazametz, 1998; Jensen et al., 1996; Li, 2003; Morris & Tersine, 1990; Shafer & Charnes, 1992, 1995; Suresh & Meredith, 1994). It also characterizes the fluidity of the material flow in the system. The *MFT* is the average time that every batch remains in the system in order to be manufactured.

Due Date Related Measures

Researchers used essentially Mean Tardiness (*MT*) and Mean Earliness (*ME*) as due date related performance measures (Farrington & Nazametz, 1998; Jensen et al., 1996). *MT* is the average over all tardy jobs of the difference between delivery date and the promised due date. *ME* is similarly obtained for all early jobs. Other researchers used the percentage of tardy jobs (*TARDY*) and the percentage of early jobs (*EARLY*).

Other Measures

FL-CL comparative studies consider several other performance measures. The system *Throughput*, considered as productivity measure, is the average number of parts exiting the system by time unit (Faizul huq et al., 2001; Morris & Tersine, 1994). It is also used for detecting the attainment of steady state indicator in a simulation run. Besides, some studies used the operator utilization rate (*OPUR*) (Morris & Tersine, 1994), the average machine utilization rate (*MUR*) (Farrington & Nazametz, 1998; Morris & Tersine, 1994; Shafer, & Charnes, 1995), the mean "queue" waiting time (Pitchuka et al., 2006) or the average *ST/PT* ratio (Li, 2003) as performance indicators. The first two measures must be maximized to ensure a high degree of resource exploitation but the third and fourth measures should be minimized to improve the efficiency of the MS.

Comparative Studies Findings

By means of four simulation experiments, Morris and Tersine (1990) examined the influence of the ratio *ST/PT, TT*, *DEMAND* stability and parts *FLOW* within cells on the performance of CLs compared to FLs. In this comparative study, the performances were measured using *MFT* and *WIP*.

Results demonstrate that in the quasi totality of the tested contexts, the FL always outperforms the CL and generates smaller *MFT* and *WIP*. Besides, comparison results reveal that the ideal context for CL must be characterized by a high *ST*/*PT* ratio, a stable *DEMAND*, a unidirectional *FLOW* and a substantial *TT* between process departments.

To find out the operating conditions under which the CL outperform the FL, Shafer and Charnes (1992) investigated 24 combinations of *DD*, *mopt*, *PT* and *BS*. As for the previous study, authors used the same performance measures. They found the CL superior to the FL in all operating conditions according to both performance measures.

Morris and Tersine (1994) extended the results of their first comparison study (Morris & Tersine, 1990) by investigating the impact of a dual resource constrained shop on the performances of CL and FL using three operator scheduling rules in the CL. Simulation observations were collected for four performance measures including the mean *Throughput*, the *WIP*, the *MUR* and the *OPUR*. Results reveal that the FL outperformed the CL on all of the used performance measures regardless of the operator scheduling rule.

Authors investigate the sensitivity of their results relatively to changes in shop congestion level and changed respectively the *IAT* and *OPUR* in two other experiments. It appeared that the FL still outperforms the CL.

Besides, Suresh and Meredith (1994) aimed to overcome the loss of pooling synergy in the CL. Hence, they used simulation in order to compare the CL to an efficiently operated FL (EFL) using average *MUR*, *WIP* and *MFT* to assess the two layout's performance measures. The EFL is characterized by an optimal *BS*, a reduced *TT* and part-family-oriented scheduling rules. The main experimental factors involved in this study were *PT*, *ST*, *BS*, δ and *IAT*. First, every experimental factor was tested separately. Then all the experimental factors were tested together. The FL was found to be superior to the CL for large batch sizes (*BS*>32). However, for relatively small *BS*, the CL could outperform the FL if δ is smaller than 0.2. Comparison results do not change when the variability of *PT*, *ST* or *IAT* were separately reduced. On the other hand, if all factor effects were combined, the CL outperformed the FL even for small *BS*.

Shafer and Charnes (1995) used simulation to study a manufacturing context inspired from Morris and Tersine (1990). In fact, they used the same levels of the following factors: *t*, *f*, *c*, *d*, *Mn*, *Mf* and *mopt*. Authors aimed to compare alternative loading procedures for CL and FL in a variety of operating environments defined by combinations of 4 factors: *FLOW*, *TT,* labor constraints and MS congestion level. The third factor was modeled using two levels of the operator number while the last factor was modeled through the variation of the *PT*. Besides, each layout was investigated using two loading policies. For the FL the first policy permitted machine dedication while the second did not. On the other hand, for CL the first policy restricted the processing to only one batch at a time in a cell and the second allowed the processing of different batches at the same time. Both policies authorized CL operations overlapping. The authors used *MFT* and *WIP* in a two stage comparison methodology. In the first stage, labor constraints were not considered. In the second stage, a constraint was imposed on labor allowing only 8 operators to the whole shop in both configurations. It is worth noting here that the presence of one operator is required during setups and processing operations. The first stage simulation results demonstrate that the two layouts were equivalent regarding *WIP* while the CL generated lower *MFT* than the FL. In contrast, in the second comparison stage the FL showed lower *MFT* than the CL. The authors justified this result by the labor constraint effect on the CL. Indeed, according to the authors the labor constraints handicaps more seriously the CL since it reduces the operations overlapping possibilities while its effect on the

FL is not significant since the departments have only 3 machines in average.

Another study by Jensen et al. (1996) assessed the FL and CL performances through *MFT*, *WIP MT*, *ME* and *TARDY*. They based their study on a full- factorial experimental plan involving layout type, *RULE*, *DEMAND* variability and δ as experimental factors. To determine the influence of each factor on the studied performance measures, the authors analyzed their simulation results by ANOVA. Aside from the layout type, the most influent factor was found to be *DEMAND* variability followed by δ and *RULE*. Then, the authors performed a pairwise comparison of *RULE*. Results demonstrate that SPT-L and EDD were the best performing rules regarding *MFT* and *MT* respectively. Finally, they compared layouts using the best found *RULE*. The results of this final step revealed that the FL was always superior to the CL with regard to all performance measures.

As for Farrington and Nazemetz (1998), their comparative study is based on a three-factor-full-factorial experimental plan. The three experimental factors were the layout type, the *PT* variability and the *IAT* variability. It's worth noting here that the high variability level was associated to a small *BS* and vice versa. They assessed the two layouts using different performance measures, namely *MFT*, *WIP*, *TARDY*, *MUR* and a number of others less common measures. Comparison results prove that the FL is superior to the CL in a context defined by a high variability of *PT* and a low variability of *IAT*. But, when both factors show high variability, the performances of the two layouts are close. Besides, The CL outperforms the FL in all remaining conditions.

Faizul huq et al. (2001) presented in their comparison study a straightforward two-factor-full-factorial simulation plan using the *MFT* and the *Throughput*. The two studied factors were *BS* andδ. For the sake of objectivity, the authors used the EFL concept. ANOVA investigation showed that the two layout *Throughput* performances were not significantly different. In deed, the two layouts presented significant differences in some of the studied combinations only in terms of *MFT*. In fact, The CL outperformed the FL only for small *BS* and very largeδ. In all other conditions, the FL was clearly superior.

Regarding Li (2003), the author used *MFT* and *WIP* to explore the superiority domains of both layouts in a diversity of contexts. These contexts are defined by the *FLOW*, the *TM*, the variability of *PT*, the variability of *ST* and finallyδ. The performance measures results analysis showed that the major factor in establishing the superiority of one of the two layouts isδ. Hence, the CL outperformed the FL at high level of δ and the FL was the best layout in the low δ region. Both layouts showed equivalent performance measures for intermediate value ofδ.

The last reviewed study, done by Pitchuka et al. (2006), compared FL to CL using a four-factor-full-factorial experimental plan featuring *PT*, *ST*, *BS* and *IAT*. The authors considered only the "queue" waiting time as performance measure. It was shown that the CL can outperform the FL in the majority of the studied contexts. Indeed, in the CL numerous work centers generated inferior "queue" times to those of the corresponding work centers in the FL.

Objectivity Assessment

Conditions Favoring FL

Jensen et al. (1996), Pitchuka et al. (2006) and Shafer and Charnes (1992) considered very low *TT* which implicitly advantage the FL, since one of the main advantages of the CL is time saving by locating machines required to manufacture a part close to each other. On the other hand, Jensen et al. (1996), Morris and Tersine, (1990, 1994) and Pitchuka et al. (2006) used a CL with no operations overlapping allowed in part processing. This does not permit to take advantage of CL benefits. Moreover, Farrington and Nazametz (1998) stated that they chose not to reduce the *ST* in the CL context. Their motivation was to avoid any biases in favor of the CL. But, by doing so, they favored the FL since they eliminated one of the main advantages of the CL.

Conditions Favoring CL

The study of Shafer and Charnes (1992) is obviously biased in favor of the CL. In deed, the authors consider single-machine departments. So, they eliminate the main and probably the only benefit of this type of layout: the pooling synergy effect between same department machines. Consequently, the results were clearly in favor of the CL even with the assumption of null transfer times advantaging the FL. Regarding Li (2003), the study featured unidirectional cell *FLOW* by duplicating the necessary machines to avoid backtracking. This indirectly led to the reduction of the cell number. The machine duplication within cells biased the comparison results in favor of the CL. In fact, this attaches to the CL the main advantages of the FL which is the synergy between functionally equivalent machines. As for Suresh and Meredith (1994), they used FL *TT* relatively very high compared to the *PT*. This probably advantage the CL and make clear why its performance are superior to the performance of the FL in almost all the testing contexts even though no operations overlapping has been used in the CL.

Other Conditions

This category essentially includes the lack of vital information about the used experimental factor settings as well as key elements defining the manufacturing contexts. Indeed, even if Morris and Tersine, (1990, 1994) provided in their studies the material handling equipment speed, they did not mention any distances between departments or machines. These distances are required in order to evaluate the *TT* in the two layouts. On the other hand, Farrington and Nazametz (1998) and Shafer and Charnes (1992) did not mention the *RULE* they used. In addition, Farrington and Nazametz (1998) failed to report numerous key experimental factors such as *mopt*, *IAT* and *PT*. Despite its established importance, Jensen et al. (1996) did not use the *BS* as an experimental factor neither did they mention its constant value used throughout the investigation.

Other lacks of important data are included in this category, particularly technical simulation-related information such as the replication length (Farrington & Nazametz, 1998; Li, 2003) and the warm-up period length (Farrington & Nazametz, 1998). On the other hand, numerous incongruities appear in different comparative studies. For example, the difference between the two shop configurations of the CL studied by Li (2003) is not clear. Indeed, in the figures illustrated by the authors, the arrows indicating the products *FLOW* show that there is no backtracking flow even in the CL with backtracking flow allowed. These incongruities are more serious in Faizul huq et al. (2001) study. Indeed, despite stating that no inter-cell moves were allowed, the authors defined the inter-cell travel time by a uniform law.

The use of inappropriate MS data appears especially in the study of Faizul huq and al. (2001). Indeed, the major flaw of this study is the definition of the manufacturing context. In fact, they used the same routings for the same product types. This generated three identical manufacturing cells. More gravely, the use of single-product families annuls any setup operation in the cell except for the initial setups. Hence, the factor δ becomes irrelevant and any results showing the importance of this factor are seriously questionable.

SIMULATION MODELS

Basic Simulation Features

FL and CL layouts simulation models are developed using the commercial simulation software

Arena 7.0 (Kelton, Sadowski, & Sadowski, 2002). This simulation tool integrates all the needed simulation functions including animation, analysis of input and output data. Every MS model consists of the four main components: manufacturing orders launching and attribute assignation, part transfer, part manufacturing and statistics collection.

Manufacturing Orders Launching and Attribute Assignation

Manufacturing orders (MOs), being batches of parts of the same type, are launched by "Create" modules. Every "Create" module defines batches *IAT* following the used probabilistic rule in addition to their *BS*. A specific "Create" module is dedicated for each part type. As soon as parts are lunched, they pass through an "Assign" module where characteristics are attributed to them. These characteristics are either time-related such as *PT* and *ST*, or also identification indicators such as part's type as well as part's family and factors necessary to the MS piloting like part's routing or "Sequence".

Part Transfer

Parts are transferred, either individually or in batches, between physical locations modeled by the "Station" modules, in which they should undergo the required manufacturing steps. These locations are either machines in CL or departments in FL. Transfer are carried out by "Route" modules permitting to prescribe destinations as well as transfer times. These modules use "Sequence" attribute of the transferred parts in order to prescribe the next destination. The "Sequence" corresponds to the part routing expressed as stations list.

Two manufacturing strategies could be followed for the parts transfer in the shops: "with operations overlapping" or "without operations overlapping". In the first strategy, parts of the same batch could be processed simultaneously on different machines of a department or a cell. In the second strategy, all parts of the same batch are processed on the same machine of the cell or department before being transferred collectively to the next machine or department. In all cases, batches must be split by "Separate" modules before accessing any machine. Batch reconstitution for transfer is performed using "Batch" modules.

Part Manufacturing

Every machine is modeled by a "Process" module, associated to a "Station" module and a "Resource" module. The "station" module determines the physical location of the machine and the "Resource" module represents the capacity and the availability of the machine itself. In fact, the "Process" module seizes the associated resource for the required period of time and then releases it. So, the machine becomes idle and available again for manufacturing another part. The machine resource is seized during a period of time that corresponds to the *PT* of the part being processed and eventually the required *ST* if the machine was set for a different part type. The *ST*, when relevant, is weighed by the setup reduction factor δ whenever the part type belongs to the family of the last processed one.

Statistics Collection

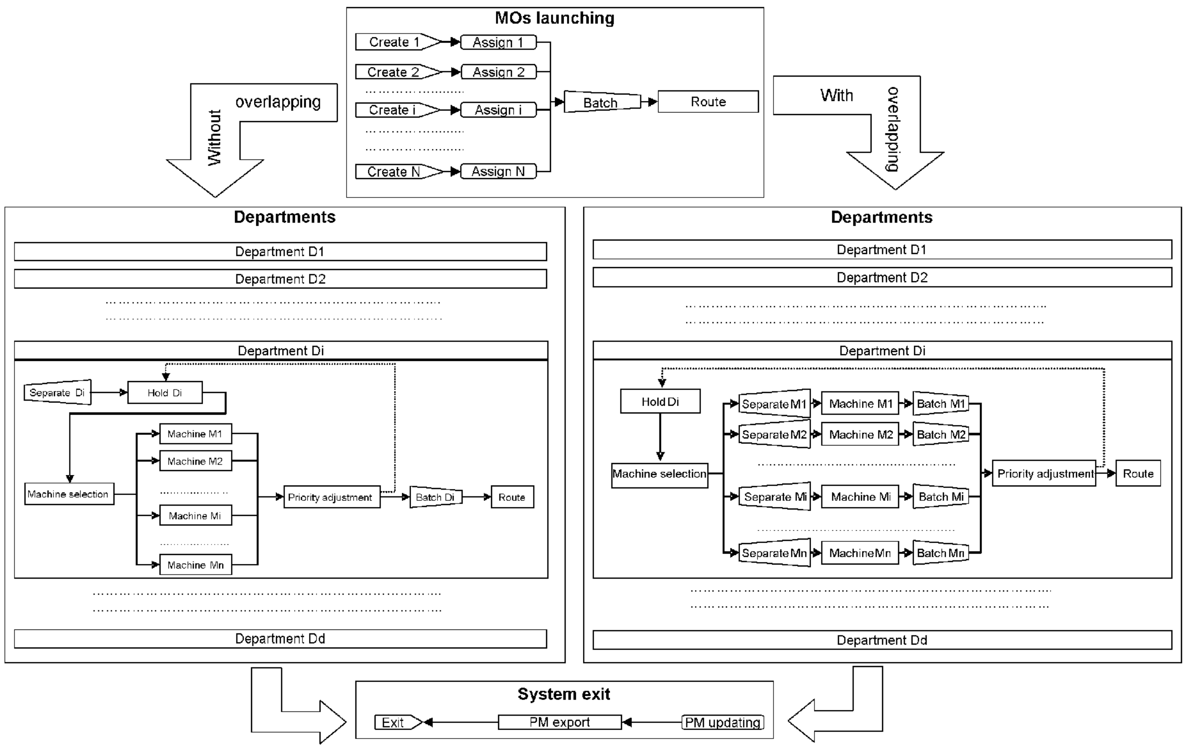
Before leaving the MS, every batch must go through an "Assign" module in which the parameters defined as performance measures are computed and updated. The acquired data is then stored in an Excel file using a "Readwrite" module for eventual treatment and analysis.

Functional Layout Model

The functional layout model is composed of three sections: "MOs launching", "Departments" and "System exit" (see Figure 1).

MOs are launched by "Create" modules dedicated each part type. Each "Create" module

Figure 1. FL model



is coupled to an "Assign" module. The generated parts are then grouped into batches and routed to their first manufacturing step's department. A batch arriving at a department is made waiting in a queue modeled by a "Hold" module. This module is governed by a priority rule that could be FCFS, SPT or any other priority rule. When at least one of the department machines becomes available, the "Hold" module releases the prioritized batch from the waiting queue. The released batch is then transferred to the "Machine selection" sub-model that selects one among the available machines. The logic of this sub-model is coherent with the waiting queue priority rule.

When operations overlapping are not allowed, every batch is split once it reaches the assigned machine. Hence, each batch can be treated only by a single machine. On the other hand, if operations overlapping are permitted, parts batches are split before accessing the department queue. So, parts become independent and could be dispatched to several machines of the same department to be processed simultaneously. In both cases, batches are gathered by a "Batch" module right after processing and before the transfer to next manufacturing step. The combination of the operations overlapping strategy, the machine selection process and the waiting queue priority rule define the shop scheduling policy.

Cellular Layout Model

The CL model is composed of "*c*" sub-models corresponding to the "*c*" MS cells. Each sub-model is composed of three sections: "MOs launching", "Machine cells" and "Cell exit".

As for the FL, MOs are launched by "Create" and "Assign" modules dedicated to each part type. The generated parts are then grouped in batches before being routed to the general cell queue. Such a queue holds part batches until their first routing step machine becomes available. In addition, each machine has its own waiting queue. Both queues are governed by the same priority rule.

If operations overlapping are allowed, batches are split just before leaving the cell general queue. Hence, every part can follow its routing without waiting for the other batch parts. Batches are finally regrouped just before the cell exit. In contrast, if operations overlapping are not implemented, every batch is split when it reaches the machine next machine on its routing. Batches are regrouped once their processing is accomplished. Then, they are transferred towards the following machine or to the system exit.

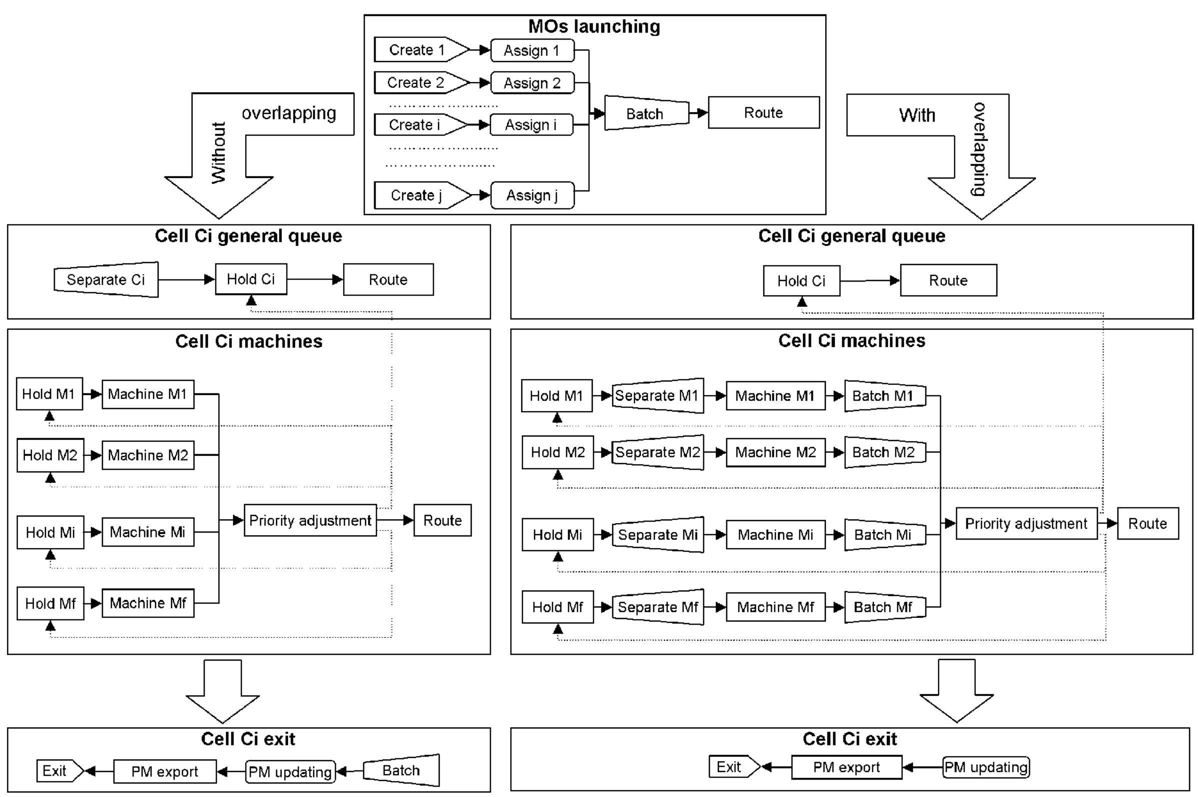
THE OBJECTIVE COMPARISON METHODOLOGY

Overview

The objective comparison methodology (OCM) aims essentially at the development of a mathematical model permitting to predict the superiority of one layout or the other. It is the product of the application of Taguchi method of experiment design. Hence, the OCM is mainly composed of 3 main phases:

- Phase 1: Choosing MS parameters and setting their levels
- Phase 2: Construction of the experiments plan, results analysis and development of the mathematical model

Figure 2. CL model



- Phase 3: Refinement of the simulation plan and improvement of the mathematical model

Each of these phases is composed of one or several stages. Some stages should be reiterated several times.

Phase 1: Choosing Levels of the Manufacturing System Parameters

In the first phase of the OCM the manager must choose the MS parameters as well as their levels. Generally, every MS can be characterized by three types of parameters: signal factors, control factors and noise factors.

Signal Factors (SF)

Signal factors are factors that are expected to affect the average response. In addition, these factors identify the manufacturing context and are kept constant in every application of the OCM. This category includes the department's number *d*, the cell's number *c*, the number of equivalent machines in every department *Mn* and the number of different machines in every cell *Mf*. The four other signal factors are the number of part families *f*, the number of part types by family *t*, the number of manufacturing operations *mopt* and the existence or no of inter-cell moves.

Control Factors (CF)

As for the signal factors, control factors can affect the average response but, more importantly, can affect the extent of the variability about the average response. These factors are to be varied throughout the simulation plan. This category includes the *ST,* the *PT*, the *TT,* the I*AT* and theδ. The three other Control factors are the *BS*, the *RULE* and the *TM*. For more objectivity of comparison results, *ST*, *TT* and *PT* are put into the following ratio forms *ST/PT* and *TT/PT*. Indeed, *ST* and *TT* being nonproductive activities, these ratios are used to compare them to *PT* which is a productive activity. In addition to the studied CFs, several factor interactions (CFI) could also be investigated in every application of the OCM. CFI between CF_x and CF_y is here noted $CF_x xCF_y$.

Noise Factors (NF)

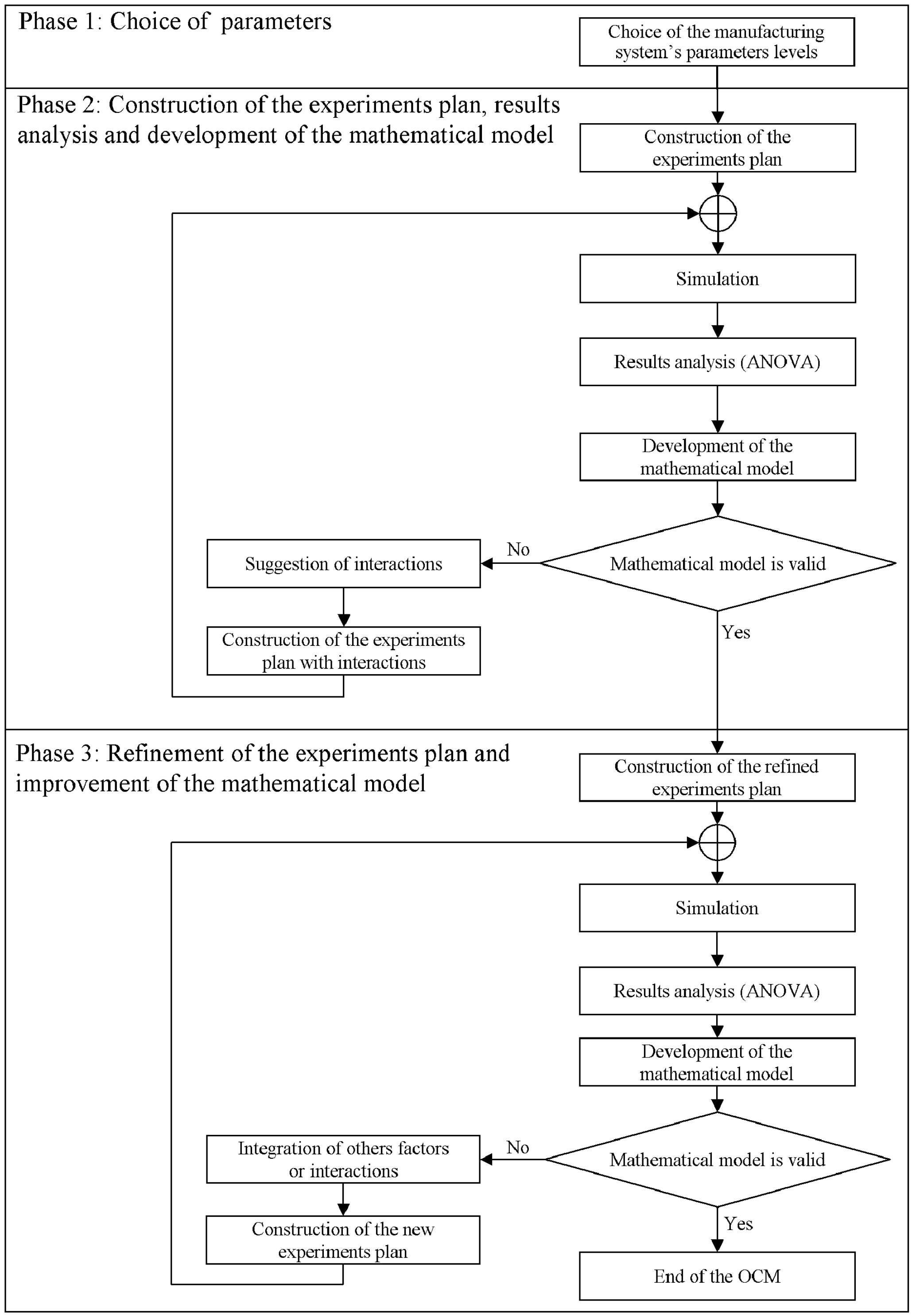
Noise factors are difficult or even impossible to control. Some of these factors could have a direct influence on the MS performances. Hence, instead of controlling them, the methodology aims at determining a solution in terms of CF that is robust relatively to unpredictable variations of *NF*.

Phase 2: Construction of the Experiments Plan, Results Analysis and Development of the Mathematical Model

The main purpose of the second phase of the OCM is to develop the mathematical model. This model gives an interpretation of the SM parameters effect's on the performances of the two layouts. It is developed through the following stages:

- **Stage 1:** An initial plan of experiments is constructed using standard OAs developed by Taguchi (Taguchi, Elsayed, & Hsiang, 1989). This plan is a set of experiments (simulations) where several CFs levels are varied from an experiment to another. It permits to considerably minimize the experimental effort.
- **Stage 2:** Simulations are conducted and performance measures of the two layouts are collected. The performance measures are expressed using the signal to noise ratio (*S/N*). This ratio is an essential indicator of the ability of the system to perform robustly in the presence of some noise effect (Park, 1998). There are three type of *S/N* ratios: lower-the-better (LB), nominal-the-

Figure 3. Overview of the OCM



best (NB), and higher-the-better (HB). In the OCM, the HB type *S/N* is used.

- L is better than FL, it is proposed to maximize the HB type S/N characterizing the MFT ratio MFTFL/MFTCL
- **Stage 3:** Simulations results are then analyzed by the analysis of variance method (ANOVA). The ANOVA establishes the relative significance of CFs in terms of their percentage contribution to the response (Phadke, 1989; Ross, 1996). The relative significance of CFs is translated by the Fischer factor "*F*" (Montgomery, 2001). The ANOVA also estimates the variance of error.
- **Stage 4:** The mathematical model is developed by interpolating the CFs effects. The validity of the developed mathematical model is then verified through the confirmation experiment. This experiment consists of adopting in an extra simulation experiment the best levels of CFs. If the average of the results of the confirmation experiment is within the limits of the confidence interval (CI) of the predicted result, then the mathematical model is considered confirmed (Kiefer, 1977). Hence the OCM can move to the following phase. Otherwise, interactions between CFs are taken in account in a new model. The second phase of the OCM is then reiterated from the third stage. This cycle should be reiterated as much as necessary to get a valid mathematical model. In each iteration, the insignificant interactions must be eliminated and replaced by other interactions.

Phase 3: Refinement of the Simulation Plan and Improvement of the Mathematical Model

The purpose of this phase is to refine the simulation plan and to improve the mathematical model developed in the second phase. So, in this plan, only the most significant CFs and CFIs are considered. Besides, for each CF, additional levels are investigated to study the non-linearity effect of the process factors. This phase is very similar to the second phase. Indeed, it essentially includes the same main stages. Only the choice of factors and interactions to integrate in the mathematical model is different. Once the improved mathematical model developed, its validity is tested.

Academic Case Study

The studied MS is inspired from the comparison study of Morris and Tersine (1990). This MS is composed by 30 machines grouped in 8 departments in the FL and 5 cells in the CL. It is also characterized by 30 part types grouped in 5 families. Every part family is composed of 6 part types. Each part type requires from 2 to 6 production operations. In addition, no inter-cell moves are required.

The FL and CL simulation models were developed using the ARENA commercial software. Observations were then collected for two performance measures: *MFT* and *Throughput*. The second measure is used solely for warm up period detection. The results show that a warm up period of 200000 minutes is needed. The models can then be run for 800000 minutes.

Choice of MS Parameters

The CFs are here studied using two levels each as depicted in Table I. It is worth noting that the original level corresponds to the level initially used in the MS.

Initial plan of Experiments

Each of the two level CFs has 1 degree of freedom (DOF). Hence, the total degree of freedom (TDOF) required for the studied seven CFs is 8 [=7×1+1]. As per Taguchi's method the total experiments number of the selected OA must

Figure 4. OCM application: Initial experiments plan

STAGES 1 & 2 : ORTHOGONAL ARRAY CONSTRUCTION AND SIMULATION RESULTS

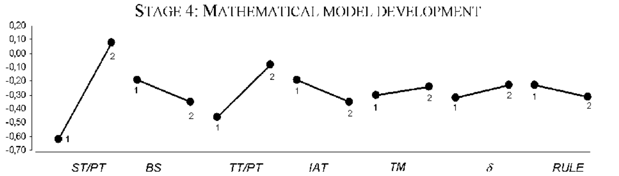
Orthogonal array								Results		
								MFT_{FL}/MFT_{CL}		S/N
Exp	$ST/PT^{(1)}$	$TT/PT^{(2)}$	$IAT^{(3)}$	$\delta^{(4)}$	$TM^{(5)}$	$RULE^{(6)}$	$BS^{(7)}$	Rep1	Rep2	
1	1	1	1	1	1	1	1	0.93	0.92	-0.69
2	1	1	1	2	2	2	2	0.93	0.90	-0.77
3	1	2	2	1	1	2	2	0.93	0.91	-0.71
4	1	2	2	2	2	1	1	0.97	0.96	-0.31
5	2	1	2	1	2	1	2	0.97	0.98	-0.24
6	2	1	2	2	1	2	1	0.98	0.98	-0.14
7	2	2	1	1	2	2	1	1.04	1.05	0.37
8	2	2	1	2	1	1	2	1.06	1.02	0.33

STAGE 3: RESULTS OF THE ANALYSIS OF VARIANCE

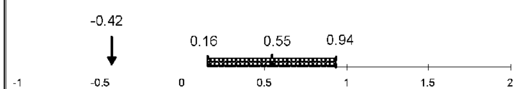
Parameter	*SS*	*DOF*	*MS*	*Pooling*	*DOFp*	*MSp*	*F*
ST/PT	0.99	1	0.99		1	0.99	73.77
BS	0.05	1	0.05		1	0.05	3.63
TT/PT	0.29	1	0.29		1	0.29	21.77
IAT	0.05	1	0.05		1	0.05	4.04
TM	0.01	1	0.01	x	–	–	–
δ	0.02	1	0.02	x	–	–	–
RULE	0.01	1	0.01	x	–	–	–
ERROR	0.00	0		0.04	3	0.01	
TOTAL	1.42	7			7		

Notes:
SS: Sum of Squared deviations *DOF*: Degree Of Freedom *MS*: Mean Square *F*: *F*-value.

STAGE 4: MATHEMATICAL MODEL DEVELOPMENT



STAGE 5: MATHEMATICAL MODEL CONFIRMATION RESULTS



be greater than or equal to the TDOF, an $L_8(2^7)$ OA was selected for the initial experiments plan (Taguchi et al., 1989). This OA has seven columns and eight experiment-runs (rows). The seven CFs are assigned to the OA columns as depicted in Figure 4 (stages1&2). Every suggested experiment by the OA is then run for 2 replications in order to compute the *S/N* ratios. Results are shown in Figure 4 (stages 1&2). The results of ANOVA indicate that only the CFs *ST/PT*, *TT/PT*, *BS* and *IAT* are statistically significant (Figure 4-stage 3). Figure 4 (stage4), that depicts the main effects of the CFs, confirms these remarks. In this figure, the importance of the CF is expressed by its slope.

Based on the computed *S/N* ratios, the mathematical model is developed by linear interpolation. In this model, every CF can take one of two values: 1 or 2, depending on the chosen parameter level:

$$S / N = -1.53 + 0.70 \times (ST / PT) - 0.16 \times BS + 0.38 \times (TT / PT) - 0.16 \times IAT + 0.07 \times TM + 0.09 \times \delta - 0.08 \times RULE \quad (1)$$

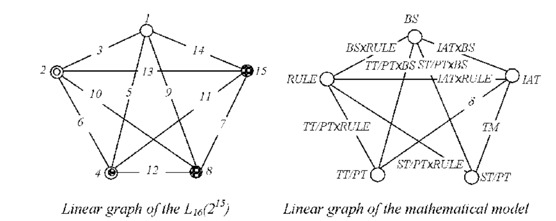
Then, the confirmation experiment considers the maximum value of *S/N* ratio to choose optimum levels of the CFs. Hence the chosen levels are ST/PT_2, TT/PT_2, IAT_1, BS_1, δ_2, TM_2 and $RULE_1$ where X_i is the ith level of the control factor *X*. In this case, the expected result in terms of *S/N* ratio is 0.55 Db. The computed 95% confidence interval is equal to CI=±0.39 Db. Therefore, the expected result should lie between 0.16 Db and 0.94 Db. As depicted in Figure 4 (stage5) the best expected response of -0.42 Db obtained by the confirmation experiment is outside the limits of the CI. The mathematical model is hence considered invalid. Additional analysis and experimentation are needed.

Simulation Plan with Interactions

Two additional iterations were needed to obtain a valid mathematical model. Only the results of the second iteration are depicted here. Based on the first iteration simulation plan ANOVA results, the simulation plan in the second iteration considers *TT/PT×RULE*, *ST/PT×RULE*, *ST/PT×BS*, *BS×RULE, TT/PT×BS, IAT×RULE* and *IAT×BS* as CFIs. Each of these CFIs has 1 DOF. The required TDOF is then equal to14 [=7×1+6×1+1]. Hence, the $L_{16}(2^{15})$ is the OA to use. This OA has fifteen columns and sixteen experiment-runs. The factors were assigned to the $L_{16}(2^{15})$ OA us-

Figure 5. OCM application: Simulation plan with interaction (Second iteration)

STAGES 1 & 2 : ORTHOGONAL ARRAY CONSTRUCTION AND SIMULATION RESULTS



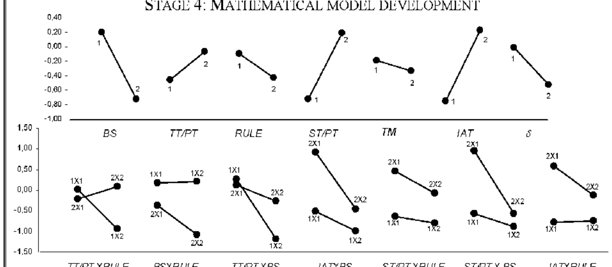
Linear graph of the $L_{16}(2^{15})$ *Linear graph of the mathematical model*

Orthogonal array								Results		
								MFT_{FL}/MFT_{CL}		S/N
Exp	$BS^{(1)}$	$RULE^{(2)}$	$TT/PT^{(4)}$	$TM^{(7)}$	$ST/PT^{(8)}$	$\delta^{(11)}$	$IAT^{(15)}$	Rep1	Rep2	
1	1	1	1	1	1	1	1	0.93	0.92	-0.69
2	1	1	1	1	2	2	2	1.26	1.25	1.97
3	1	1	2	2	1	1	2	0.98	0.99	-0.13
4	1	1	2	2	2	2	1	0.94	0.96	-0.42
5	1	2	1	2	1	2	2	0.95	0.97	-0.33
6	1	2	1	2	2	1	1	1.02	1.01	0.14
7	1	2	2	1	1	2	1	0.88	0.88	-1.09
8	1	2	2	1	2	1	2	1.26	1.30	2.14
9	2	1	1	2	1	2	2	0.97	0.93	-0.47
10	2	1	1	2	2	1	1	0.91	0.94	-0.71
11	2	1	2	1	1	2	1	0.84	0.89	-1.28
12	2	1	2	1	2	1	2	1.11	1.12	0.97
13	2	2	1	1	1	1	1	0.85	0.83	-1.52
14	2	2	1	1	2	2	2	0.86	0.73	-2.06
15	2	2	2	2	1	1	2	0.98	0.97	-0.25
16	2	2	2	2	2	2	1	0.95	0.95	-0.47

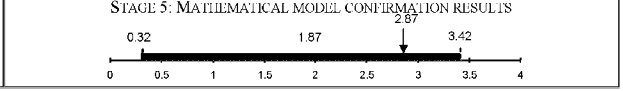
STAGE 3: RESULTS OF THE ANALYSIS OF VARIANCE

Parameter	*SS*	*DOF*	*MS*	*Pooling*	*DOFp*	*MSp*	*F*
BS	3.40	1	3.40		1	3.40	7.54
TT/PT	0.62	1	0.62	x	–	–	–
RULE	0.45	1	0.45	x	–	–	–
ST/PT	3.33	1	3.33		1	3.33	7.40
TM	0.07	1	0.07	x	–	–	–
IAT	3.87	1	3.87		1	3.87	8.59
δ	1.05	1	1.05		1	1.05	2.33
TT/PTxRULE	1.59	1	1.59		1	1.59	3.52
BSxRULE	0.54	1	0.54	x	–	–	–
TT/PTxBS	1.17	1	1.17		1	1.17	2.59
IATxBS	0.79	1	0.79	x	–	–	–
ST/PTxRULE	0.13	1	0.13	x	–	–	–
ST/PTxBS	1.45	1	1.45		1	1.45	3.22
IATxRULE	0.57	1	0.57	x	–	–	–
ERROR	0.43	1	0.43	3.17	8	0.45	
TOTAL	19.46	15			15		

STAGE 4: MATHEMATICAL MODEL DEVELOPMENT



STAGE 5: MATHEMATICAL MODEL CONFIRMATION RESULTS



ing the linear graphs displayed in the Figure 5 (stages1&2). This figure also shows the associated simulation results.

ANOVA results indicate that only the CFs *BS*, *ST/PT*, *IAT* and δ are statistically significant (Figure 5- stage 3). It also demonstrates that only the CFIs $TT/PT \times RULE$, $TT/PT \times BS$ and $ST/PT \times BS$ are statistically significant. Figure 5 (stage4) illustrates the main effects of the CFs and CFIs. In this figure the importance of a CFI is expressed by the slope difference between the interaction two curves. The mathematical model is then developed:

$$\begin{aligned} S / N = & -4.44 + 1.70 \times BS - 3.12 \times (TT / PT) \\ & + 0.54 \times RULE + 3.26 \times (ST / PT) - 0.14 \\ & \times TM + 3.45 \times IAT - 0.51 \times \delta + 1.26 \\ & \times (TT / PT) \times RULE - 0.73 \times BS \times RULE \\ & + 1.08 \times (TT / PT) \times BS - 0.89 \times IAT \\ & \times BS - 0.36 \times (ST / PT) \times RULE - 1.20 \\ & \times (ST / PT) \times BS - 0.75 \times IAT \times RULE \end{aligned} \quad (2)$$

The levels of the CFs in the confirmation experiment are as follows: BS_1, TT/PT_2, $RULE_1$, ST/PT_2, TM_1, IAT_2, δ_1. Two confirmation trials were conducted and results show that the developed mathematical model is valid (Figure 5-stage5).

Refinement of the Simulation Plan and Improvement of the Mathematical Model

The refined simulation plan considers the control factors *BS*, *ST/PT*, *IAT* and δ in addition to the CFI $ST/PT \times BS$. In addition to the two studied levels, each of the three CFs was analyzed by way of a third level. This additional level corresponds to the original level as depicted in Table 1. Hence, the required TDOF is 13 [=4×2+1×4+1]. So, the $L_{27}(3^{13})$ OA was selected for the refined simulation plan and the CFs were assigned to this array using the linear graphs displayed in the Figure 6 (stages1&2). This figure depicts also the OA and

Table 1. Control factors

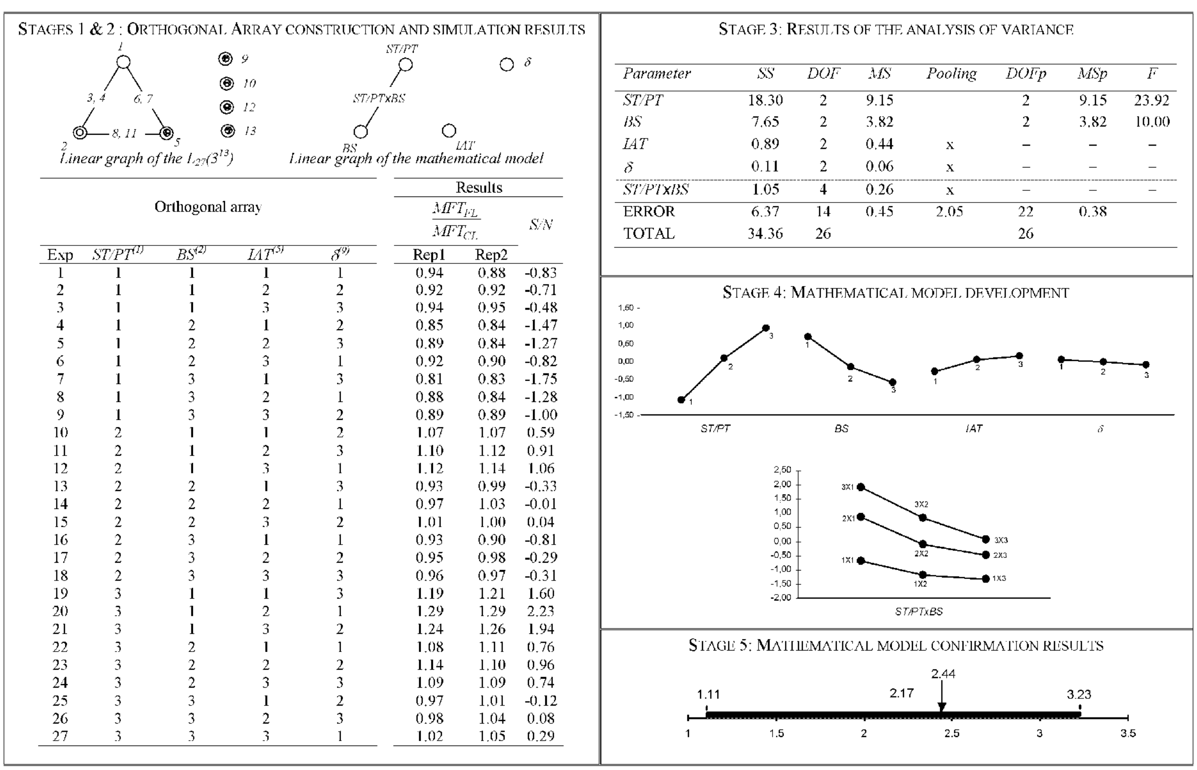
CF	Original Level	Level 1	Level 2
ST/PT	3	1	5
IAT	Exp (525) mn	Exp (420) mn	Exp (630) mn
δ	0.35	0.2	0.5
BS	38	25	50
RULE	*RL*	*RL*	*FCFS*
TM	With operations overlapping	With operations overlapping	Without operations overlapping
TT/PT	0.8 for FL ; 0.3 for CL	0.4 for FL ; 0.15 for CL	1.2 for FL ; 0.45 for CL

the simulation results. It's worth noting that the original levels of the unused CFs in the refined plan (*RULE*, *TM* and *TT/PT)* are chosen (Table 1).

The analysis of simulation results shows that only *ST/PT* and *BS* are significant (Figure 6-stage3). This observation is confirmed by the Figure 6 (stage4) that illustrates the main effects of the CFs and CFIs. The developed mathematical model is written as follows:

$$S / N = -3.83 - 0.59 \times (ST / PT)^2 - 0.15 \times BS^2 - 0.12 \times IAT^2 - 0.01 \times \delta^2 - 0.11 \times (ST / PT)^2 \times BS^2 + 0.47 \times (ST / PT)^2 \times BS + 0.43 \times (ST / PT) \times BS^2 - 2.17 \times (ST / PT) \times BS + 3.94 \times (ST / PT) + 0.65 \times BS + 0.68 \times IAT - 0.06 \times \delta \quad (3)$$

Figure 6. OCM application: Refined simulation plan



Exp	*ST/PT*(1)	*BS*(2)	*IAT*(5)	δ(9)	MFT_{FL}/MFT_{CL} Rep1	MFT_{FL}/MFT_{CL} Rep2	*S/N*
1	1	1	1	1	0.94	0.88	-0.83
2	1	1	2	2	0.92	0.92	-0.71
3	1	1	3	3	0.94	0.95	-0.48
4	1	2	1	2	0.85	0.84	-1.47
5	1	2	2	3	0.89	0.84	-1.27
6	1	2	3	1	0.92	0.90	-0.82
7	1	3	1	3	0.81	0.83	-1.75
8	1	3	2	1	0.88	0.84	-1.28
9	1	3	3	2	0.89	0.89	-1.00
10	2	1	1	2	1.07	1.07	0.59
11	2	1	2	3	1.10	1.12	0.91
12	2	1	3	1	1.12	1.14	1.06
13	2	2	1	3	0.93	0.99	-0.33
14	2	2	2	1	0.97	1.03	-0.01
15	2	2	3	2	1.01	1.00	0.04
16	2	3	1	1	0.93	0.90	-0.81
17	2	3	2	2	0.95	0.98	-0.29
18	2	3	3	3	0.96	0.97	-0.31
19	3	1	1	3	1.19	1.21	1.60
20	3	1	2	1	1.29	1.29	2.23
21	3	1	3	2	1.24	1.26	1.94
22	3	2	1	1	1.08	1.11	0.76
23	3	2	2	2	1.14	1.10	0.96
24	3	2	3	3	1.09	1.09	0.74
25	3	3	1	2	0.97	1.01	-0.12
26	3	3	2	3	0.98	1.04	0.08
27	3	3	3	1	1.02	1.05	0.29

Parameter	*SS*	*DOF*	*MS*	*Pooling*	*DOFp*	*MSp*	*F*
ST/PT	18.30	2	9.15		2	9.15	23.92
BS	7.65	2	3.82		2	3.82	10.00
IAT	0.89	2	0.44	x	–	–	–
δ	0.11	2	0.06	x	–	–	–
ST/PTxBS	1.05	4	0.26	x	–	–	–
ERROR	6.37	14	0.45	2.05	22	0.38	
TOTAL	34.36	26			26		

The confirmation experiment shows that the developed mathematical model is valid (Figure 6-stage5).

The MS manager can use this mathematical model to determine the best layout of its MS machines. He can also investigate the effect of the change of one or several CFs levels on performances of the two layouts. In fact, if the computed *S/N* ratio value is negative then the FL is the outperforming layout. In contrary, if the predicted *S/N* ratio value is positive then the CL outperforms the FL. Finally, the two layouts performances are considered equivalents if the *S/N* ratio value predicted by the mathematical model is close to zero. Table 2 depicts the CL and FL superiority contexts expressed as combinations of the CFs.

This table can be used by the manager to determine the more effective layout for every one of the 81 possible level combinations of the four considered CFs. In deed, the intersection between the line that represents the combination of the *ST/PT* and *BS* levels and the column corresponding to the *IAT* and δ levels gives the best performing layout. For example, the CL is the best layout for the following CFs levels combination: ST/PT_2, BS_1, IAT_2 and δ_2.

The mathematical model can also be used to predict the best layout for "intermediate levels" of the CFs *ST/PT*, *BS*, *IAT* andδ. Indeed unlike the *TM* and *RULE* CFs which are discrete and can be investigated only for specified levels, *ST/PT*, *BS*, *IAT* andδ are continuous factors. For example for the following setting combination: $ST/PT_{1.5}$, $BS_{2.2}$, $IAT_{1.4}$ and $\delta_{2.4}$ the FL outperforms the CL. In this case, le level X_i of the CF X is obtained by linear interpolation between the different levels of this CF.

CONCLUSION

This chapter presents an objective methodology for comparing functional and cellular layouts. This methodology aims to help MS managers choosing the appropriate layout for their manufacturing system. The developed methodology is based on the Taguchi method for the design of experiments and results analysis combined to discrete event simulation. This method permits, through a minimal experimental effort, to reliably evaluate the effect of each MS parameters on the system performances. It also reveals the possible interactions between MS parameters. The goal of this methodology is the development of a mathematical model predicting the superiority of one of the two layouts. In fact, once developed and validated, the mathematical model can be used by

Table 2. Level combinations giving layout superiority

		IAT_1			IAT_2			IAT_3		
		δ_1	δ_2	δ_3	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
ST/PT_1	BS_1	FL	FL	FL	FL	FL	FL	FL	FL	FL
	BS_2	FL	FL	FL	FL	FL	FL	FL	FL	FL
	BS_3	FL	FL	FL	FL	FL	FL	FL	FL	FL
ST/PT_2	BS_1	CL	CL	CL	CL	CL	CL	CL	CL	CL
	BS_2	FL	FL	FL	CL	FL&CL	FL	CL	CL	FL&CL
	BS_3	FL	FL	FL	FL	FL	FL	FL	FL	FL
ST/PT_3	BS_1	CL	CL	CL	CL	CL	CL	CL	CL	CL
	BS_2	CL	CL	CL	CL	CL	CL	CL	CL	CL
	BS_3	FL	FL	FL	CL	CL	CL	CL	CL	CL

the MS manager to predict the *S/N* ratios for any combination of the MS parameters. The sign of the predicted *S/N* ratio indicates the best layout. The model can also be exploited to interpolate the results between the studied levels of continuous parameters such as batch inter arrival time or batch size. An academic case study showed the capacity of this methodology for choosing the best layout for a MS.

The developed methodology can find direct applications in the industry. However, many aspects of the comparison methodology should undergo further developments. The first task is the enlargement of the application scope to other control factors such as various levels of the number of operators or different degrees of the operator's qualification. In addition, in order to minimize the effort provided by the MS manager, the automation of coupling between the simulator and the analyze software is also projected. This should increase the chance of the proposed methodology to be successfully applied and validated on real cases.

REFERENCES

Agarwal, A., & Sarkis, J. (1998). A review and analysis of comparative performance studies on functional and cellular manufacturing layouts. *Computers & Industrial Engineering*, *34*(1), 77–89. doi:10.1016/S0360-8352(97)00152-6

Faizul huq, Douglas, A. H., & Zubair, M.M. (2001). A simulation analysis of factors influencing the flow time and through-put performance of functional and cellular layouts. *Integrated Manufacturing Systems, 12* (4), 285-295.

Farrington, P. A., & Nazametz, J. W. (1998). Evaluation of the performance domain of cellular and functional layouts. *Computers & Industrial Engineering*, *34*(1), 91–101. doi:10.1016/S0360-8352(97)00153-8

Jensen, J. B., Malhotra, M. K., & Philipoom, P. R. (1996). Machine dedication and process flexibility in a group technology environment. *Journal of Operations Management*, *14*(1), 19–39. doi:10.1016/0272-6963(95)00030-5

Kelton, W. D., Sadowski, R. P., & Sadowski, D. A. (2002). *Simulation with Arena* (2nd ed.). New York, NY: McGraw-Hill.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *Journal of the American Statistical Association*, *72*, 789–827. doi:10.2307/2286460

Li, J. (2003). Improving the performance of job shop manufacturing with demand-pull production control by reducing set-up/processing time variability. *International Journal of Production Economics*, *84*(3), 255–270. doi:10.1016/S0925-5273(02)00447-4

Montgomery, D. C. (2001). *Design and analysis of experiments* (5th ed.). New York, NY: Wiley.

Morris, J. S., & Tersine, R. J. (1990). A simulation analyses of factors influencing the attractiveness of group technology cellular layouts. *Management Science*, *36*(12), 1567–1578. doi:10.1287/mnsc.36.12.1567

Morris, J. S., & Tersine, R. J. (1994). A simulation comparison of process and cellular layouts in a dual resource constrained environment. *Computers & Industrial Engineering*, *26*(4), 733–741. doi:10.1016/0360-8352(94)90008-6

Park, S. H. (1998). *Robust design and analysis for quality engineering*. London, UK: Chapman & Hall Publication.

Phadke, M. S. (1989). *Quality engineering using robust design*. Englewood Cliffs, NJ: P.T.R. Prentice Hall.

Pitchuka, L. N., Adil, G. K., & Ananthakumar, U. (2006). Effect of the conversion of the functional layout to a cellular layout on the queue time performance: Some new insights. *International Journal of Advanced Manufacturing Technology*, *31*(5-6), 594–601. doi:10.1007/s00170-005-0219-2

Ross, P. J. (1996). *Taguchi techniques for quality engineering*. New York, NY: McGraw-Hill.

Shafer, S. M., & Charnes, J. M. (1992). Cellular versus functional layouts under a variety of shop operating conditions. *Decision Sciences*, *24*(3), 665–681. doi:10.1111/j.1540-5915.1993.tb01297.x

Shafer, S. M., & Charnes, J. M. (1995). A simulation analyses of factors influencing loading practices in cellular manufacturing. *International Journal of Production Research*, *33*(1), 279–290. doi:10.1080/00207549508930148

Shambu, G., Suresh, N. C., & Pegels, C. C. (1996). Performance evaluation of cellular manufacturing systems: A taxonomy and review of research. *International Journal of Operations & Production Management*, *16*(8), 81–103. doi:10.1108/01443579610125796

Suresh, N. C., & Meredith, J. R. (1994). Coping with the loss of pooling synergy in cellular manufacturing systems. *Management Science*, *40*(4), 466–483. doi:10.1287/mnsc.40.4.466

Taguchi, G., Elsayed, E., & Hsiang, T. (1989). *Quality engineering in production systems*. New York, NY: McGraw-Hill.

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 189-206, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 90
Random Dynamical Network Automata for Nanoelectronics: A Robustness and Learning Perspective[1]

Christof Teuscher
Portland State University, USA

Thimo Rohlf
Genopole, France

Natali Gulbahce
Northeastern University, USA

Alireza Goudarzi
Portland State University, USA

ABSTRACT

It is generally expected that future and emerging nanoscale computing devices will be built in a bottom-up way from vast numbers of simple, densely arranged components that exhibit high failure rates, are relatively slow, and connected in an unstructured way. Other than that, there is little to no consensus on what type of technology and computing architecture holds most promises to go far beyond today's top-down engineered silicon devices. Highly structured crossbar-like and cellular automata architectures have been proposed as possible alternatives to the von Neumann computing architecture, which is not generally well suited for emerging, massively parallel and fine-grained nanoscale electronics. While the top-down engineered semi-conducting technology favors regular and locally interconnected structures, emerging bottom-up self-assembled devices tend to have to be unstructured and heterogeneous because of the current lack of precise control over these processes. In this paper, we survey and assess two types of random dynamical networks, namely Random Boolean Networks (RBNs) and Random Threshold Networks (RTNs), as candidates for alternative computing architectures and models for future nanoscale information processing devices. In a high-level approach that is based on previous work, we illustrate that they have the potential to offer superior properties over highly structured crossbar- or mesh-like cellular automata architectures, such as an inherent and scale-invariant robustness, more efficient communication capabilities, manufacturing benefits for bottom-up self-assembled devices, and the ability to learn and solve tasks successfully. We also show that RBNs can learn and generalize. Our investigation is driven by the need for alternative computing and manufacturing paradigms to mitigate some of the challenges traditional approaches face.

DOI: 10.4018/978-1-4666-1945-6.ch090

INTRODUCTION AND MOTIVATION

The advent of multicore architectures and the slowdown of the processor's operating frequency increase are signs that CMOS miniaturization is increasingly hitting fundamental physical limits. A key question the industry faces is how computing architectures will evolve as we reach these fundamental limits. What type of device and architecture will guarantee a sustainable and continuous progress in the information and communication technology for the next 10-20 years? A likely possibility within the realm of CMOS technology is that the integration density will cease to increase at some point, instead, only the number of components, i.e, the number of transistors, will continue to increase, which will lead to chips with a higher area. This trend can be observed with current multi-core architectures, and is expected to continue. Besides the unsolved challenge of programming large multi-core systems, the trend has implications on the interconnect architecture, the power consumption and heat dissipation, and the reliability. Another possibility lies in the smooth and incremental transition to hybrid systems, which combine traditional silicon technology with new nano- and molecular-scale components, such as carbon nanotubes and nanowires. Yet another possibility is to go beyond silicon-based technology and to radically change the computing and manufacturing paradigms, by using for example bottom-up self-assembled devices. Self-assembling nanowires (Ferry, 2008) or carbon nanotube electronics (Avouris, Chen, & Perebeinos, 2007) are promising candidates, although none of them has resulted in electronics that is able to compete with traditional CMOS so far. What seems clear is that the current way with build computers and the way we algorithmically solve problems with them may need to be fundamentally revisited. The goal of this paper is to explore such a radical new approach and to evaluate its potential.

While the top-down engineered CMOS technology favors regular and locally interconnected structures, future bottom-up self-assembled devices tend to have irregular structures because of the current lack of precise control over these processes. We therefore hypothesize that, compared to current CMOS technology, which allowed engineers to implement a logic-based computing architecture with extreme precision and reliability, future and emerging computing architectures will be increasingly driven by manufacturing constraints and particularities. Independent of the emerging device and fabrication technologies, we assume in this paper that future nanoscale devices will be built from (1) vast numbers of densely arranged devices that (2) are arranged and interconnected in some unstructured way and that (3) exhibit high failure rates. We take this working hypothesis for granted in this paper and address it from a perspective that focuses on the interconnect topology. This is justified by the fact that the importance of interconnects on electronic chips has outrun the importance of transistors as a dominant factor of performance (Meindl, 2003; Ho, Mai, & Horowitz, 2001; Davis et al., 2001). The reasons for that are twofold: (1) the transistor switching speed for traditional silicon is much faster than the average wire delays and (2) the required chip area for interconnects has increased.

In (Zhirnov, Cavin, Leeming, & Galatsis, 2008), Zhirnov et al. explored integrated digital *Cellular Automata* (CA) architectures—which are highly regular structures with local interconnects (see Section)—as an alternative paradigm to the von Neumann computer architecture for future and emerging information processing devices. They argue that CAs are well suited for semiconductor technology, which "[…] favors the realization of regular, locally connected structures […]." Here, we are interested to explore and assess a more general class of discrete dynamical systems, namely *Random Boolean Networks* (RBNs) and *Random Threshold Networks* (RTNs), for emerging technologies. We will focus on RBNs, but RTNs are included in this paper because they offer an alternative paradigm to Boolean logic, which offers significant potential for new and efficient devices.

We are interested to provide answers to the following questions:

- Do RBNs and RTNs offer benefits over CA-architectures? If yes, what are they?
- How does the interconnect complexity compare between RBNs/RTNs and CAs?
- Does any of these architectures allow to solve certain problems more efficiently?
- Is any of these architectures inherently more robust to simple errors?
- Can CMOS and beyond-CMOS devices provide a benefit for the fabrication of any of these architectures?

We will argue and illustrate that—at least from a theoretical perspective—random dynamical networks offer superior properties over classical regular CA-based architectures, such as inherent robustness as the system scales up, more efficient information communication capabilities, and manufacturing benefits for bottom-up fabricated devices, which motivates this investigation. We will present recent results on the dynamic behavior and robustness of such random dynamical networks while also including manufacturing issues in this mini-assessment.

To answer the above questions, we will present and extend recent results on the complex dynamical behavior of discrete random dynamical networks (Rohlf, Gulbahce, & Teuscher, 2007), their ability to solve problems (Mesot & Teuscher, 2005; Teuscher, Gulbahce, & Rohlf, 2007), novel interconnect paradigms (Teuscher, 2007; Teuscher & Hansson, 2008), damage spreading in spatial and small-world networks (Lu & Teuscher, 2009), and self-assembled interconnect models (Teuscher et al., 2009).

RANDOM DYNAMICAL NETWORKS

Random Boolean Networks

A *Random Boolean Network* (RBN) (Kauffman, 1968; Kauffman, 1984; Kauffman, 1993) is a discrete dynamical system composed of N nodes, also called *automata*, *elements* or *cells*. Each automaton is a Boolean variable with two possible states: $\{0,1\}$, and the dynamics is such that

$$F : \{0,1\}^N \mapsto \{0,1\}^N , \tag{1}$$

where, $F = (f_1, \ldots f_i, \ldots f_N)$, and each f_i is represented by a look-up table of inputs randomly chosen from the set of N nodes. Initially, K_i neighbors and a look-table are assigned to each node at random. Note K_i (i.e., the fan-in) can refer to the *exact* or to the *average* number of incoming connections per node.

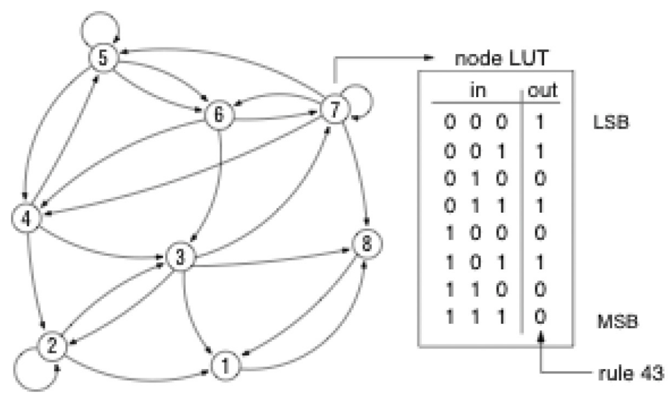
A node state $\sigma_i^t \in \{0,1\}$ is updated using its corresponding Boolean function:

$$\sigma_i^{t+1} = f_i(x_{x_{i_1}}^t, x_{x_{i_2}}^t, \ldots, x_{x_{K_i}}^t) \tag{2}$$

These Boolean functions are commonly represented by *lookup-tables* (LUTs), which associate a 1-bit output (the node's future state) to each possible K-bit input configuration. The table's out-column is called the *rule* of the node. Note that even though the LUTs of a RBN map well on an FPGA or other memory-based architectures, the random interconnect in general does not.

We randomly initialize the states of the nodes (initial condition of the RBN). The nodes are updated synchronously using their corresponding Boolean functions. Other updating schemes exist, see for example (Gershenson, 2003) for an overview. Synchronous random Boolean networks as introduced by Kauffman are commonly called *NK* networks or models. Figure 1 shows a pos-

Figure 1. Illustration of a random Boolean network with N=8 nodes and K=3 inputs per node (self-connections are allowed). The node rules are commonly represented by lookup-tables (LUTs), which associate a 1-bit output (the node's future state) to each possible K-bit input configuration. The table's out-column is commonly called the rule of the node



sible *NK* random Boolean network representation (N=8, K=3).

Random Threshold Networks

Random Threshold Networks (RTNs) are another type of discrete dynamical systems. An RTN consists of N randomly interconnected binary sites (spins) with states $\sigma_i = \pm 1$. For each site i, its state at time t+1 is a function of the inputs it receives from other spins at time t:

$$\sigma(t+1) = \text{sgn}(f_i(t)) \tag{3}$$

with

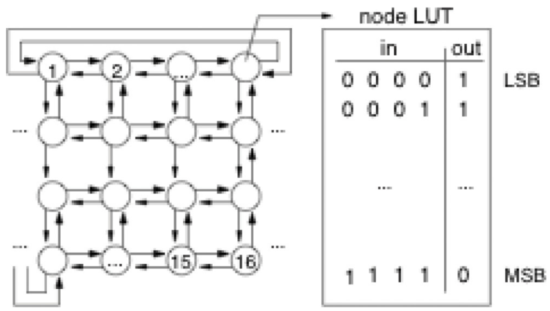
$$f_i(t) = \sum_{j=1}^{N} c_{ij}\sigma_j(t) + h \tag{4}$$

The N network sites are updated synchronously. In the following, the threshold parameter h is set to zero. The interaction weights c_{ij} take discrete values c_{ij}=+1 or ↕1 with equal probability. If i does not receive signals from j, one has c_{ij}=0.

CELLULAR AUTOMATA ARCHITECTURES

Cellular automata (CA) (Wolfram, 1984) were originally conceived by Ulam and von Neumann (Neumann, 1966) in the 1940s to provide a formal framework for investigating the behavior of complex, extended systems. CAs are a special case of the more general class of random dynamical networks, in which space and time are discrete. A CA usually consists of a D-dimensional regular lattice of N lattice sites, commonly called *nodes*, *cells*, *elements*, or *automata*. Each cell i can be in one of a finite number of S possible states and further consists of a transition function (also called *rule*), which maps the neighboring states to the set of cell states. CAs are called *uniform* if all cells contain the same rule, otherwise they are *non-uniform*. Each cell takes as input the states of the cells within some finite local neighborhood. Here, we only consider non-uniform, two-dimensional (D=2), folded, and binary CAs (S=2) with a radius-1 von Neumann neighborhood, where each cell is connected to each of its four immediate neighbors only. Figure 2 illustrates such an CA. The Boolean functions in each node must therefore define 2^4=16 possible input combinations. To

Figure 2. Illustration of a binary, 2D, folded cellular automaton with N=16 cells. Each node is connected to its four immediate neighbors (von Neumann neighborhood)



be able to compare CAs with RBNs, we do not consider self-connections.

DAMAGE SPREADING AND CRITICALITY

Random Boolean and Threshold Networks

As we have seen above, RBNs and their complex dynamic behavior are essentially characterized by the average number of incoming links (fan-in) per node (e.g., Figure 1 shows a *K*=3 network with 3 incoming links per node). It turns out that in the thermodynamic limit, i.e., $N \to \infty$, RBNs exhibit a dynamical order-disorder transition at a sparse critical connectivity (Derrida & Pomeau, 1986) (i.e., where each node receives on average two incoming connections from two randomly chosen other nodes), which partitions their operating space into 3 different regimes: (1), sub-critical, where $\left\langle K \right\rangle < K_c$, (2) complex, where $\left\langle K \right\rangle = K_c$, and (3) supercritical, where $\left\langle K \right\rangle > K_c$. In the sub-critical regime, the network dynamics are too "rigid" and the information processing capabilities are thus hindered, whereas in the supercritical regime, their behavior becomes chaotic. The complex regime is also commonly called the "edge of chaos," because it represents the network connectivity where information processing is "optimal" and where a small number of stable attractors exist.

Similar observations were made for sparsely connected random threshold (neural) networks (RTN) (Rohlf & Bornholdt, 2002) for $K_c = 1.849$. For a finite system size *N*, the dynamics of both systems converge to periodic attractors after a finite number of updates. At K_c, the phase space structure in terms of attractor periods (Albert & Barabási, 2000), the number of different attractors (Samuelsson & Troein, 2003) and the distribution of basins of attraction (Bastolla & Parisi, 1998) is complex, showing many properties reminiscent of biological networks (Kauffman, 1993).

RESULTS

In (Rohlf et al., 2007) we have systematically studied and compared damage spreading (i.e., how a perturbed node-state influences the rest of the network nodes over time) at the sparse percolation (SP) limit for random Boolean and threshold networks with perturbations. In the SP limit, the damage induced in a network (i.e., by changing the state of a node) does not scale with system size. Obviously, this limit is relevant to information and damage propagation in many technological and natural networks, such as the Internet, disease spreading in populations, failure propagation in power grids, and networks-on-chips. We measure the damage spreading by the following methodology: the state of one randomly chosen node is changed. The damage is then measured as the Hamming distance between a damaged and undamaged network instance after a large number of *T* system updates. For an electronic system, for example, we can compare such a damage event with a single event upset caused by electromagnetic radiation.

We have shown that there is a characteristic average connectivity $K_s^{RBN} = 1.875$ for RBNs and $K_s^{RTN} = 1.729$ for RTNs, where the damage spreading of a single one-bit perturbation of a network node remains constant as the system size *N* scales up. Figure 3 illustrates this newly discovered point for RBNs and RTNs. For more details, see (Rohlf et al., 2007).

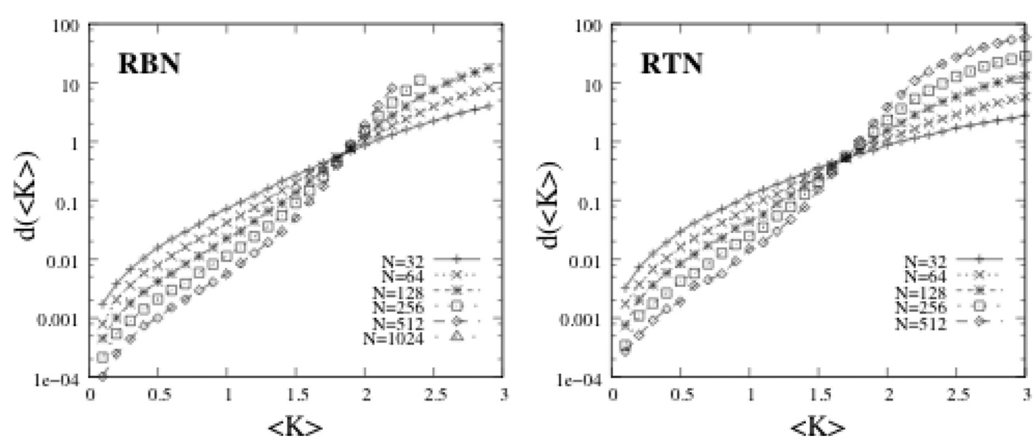
DISCUSSION

Both and are highly relevant for nano-scale electronics for the following reason: assuming we can build massive numbers of *N* simple logic gates that implement a random Boolean function, the above findings tell us that on average, every gate should be connected somewhere close to both and in order to (1) guarantee optimal robustness against failures for any system size and (2) optimal information processing at the "edge of chaos." We are also hypothesizing that natural systems, such as the brain or genetic regulatory networks, may have evolved towards these characteristic connectivities. This remains, however, to be proved and is part of ongoing research.

Cellular Automata

We have used the same approach as described above to measure the damage spreading in cellular automata. In order to vary the average number $\langle K \rangle$ of incoming links per cell in a cellular automata (e.g., as pictured in Figure 2), we have adopted the following methodology: (1) for a desired average number of links per cell $\langle K \rangle$ for a given CA size of *N* cells, the total number of links in the automaton is given by $L=N\langle K \rangle$; (2) we then randomly choose *L* possible connec-

Figure 3. Average Hamming distance (damage) $\langle d \rangle$ after 200 system updates, averaged over 10,000 randomly generated networks for each value of $\langle K \rangle$, with 100 different random initial conditions and one-bit perturbed neighbor configurations for each network. For both RBN and RTN, all curves for different N approximately intersect in a characteristic point K_s



tions on the regular CA-grid with uniform probability and establish the links. Damage is induced in the same way as for RBNs and RTNs: the state of one (or several) randomly chosen node(s) is changed. The damage is measured as the Hamming distance between a damaged and undamaged CA instance after a large number of T system updates, in our case T=200.

RESULTS

Figures 4, 5, and 6 show the average damage of both RBNs and CAs for different system sizes and for a damage size of 1 and 10 respectively. We have left out RTNs for this analysis. As one can see, both the RBN and the CA average damage for different N approximately intersect in the characteristic point $K_s^{RBN} = 1.875$. This point is less pronounced for the larger damage sizes (Figures 5 and 6). The reasons for this are finite size effects because the damage becomes too big compared to the system size. This can be seen in Figure 6 for the RBN *N=100* curve, which represents an outlier, i.e., does not intersect well with the other curves. The RBN curves confirm what was already shown above in Figure 3, and are merely plotted here for comparison with the CA architectures and their system sizes imposed by square lattices.

Interestingly, the CAs show different damage propagation behavior for different system sizes and connectivities. First, we observe that the average damage for one-bit damage events (Figure 4) is independent of the system size N for up to approximately $\langle K \rangle$=2.5 average incoming connections per cell. This behavior disappears completely for large damage sizes (Figure 6). Second, Figure 4 shows that all curves intersect at $K_s^{RBN} = K_s^{CA} = 1.875$. Third, Figure 6 suggest that for larger damage sizes, K_s^{CA} disappears for CAs. Fourth, the average damage for larger damage events, i.e., 10 and 20 in our examples, converges to the same final values for both RBNs and CAs as $\langle K \rangle$ approaches 4.

Figure 4. Average Hamming distance (damage) $\langle d \rangle$ after 200 system updates, averaged over 100 randomly generated networks for each value of $\langle K \rangle$, with 100 different random initial conditions and a damage size of 1 node for each network. See text for discussion

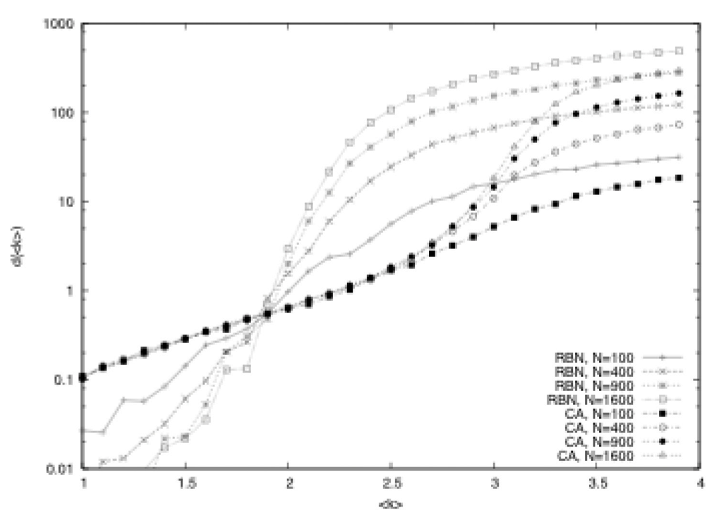
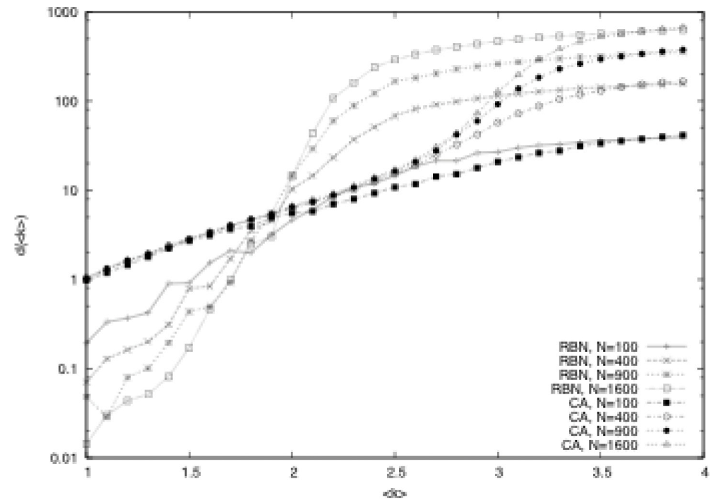


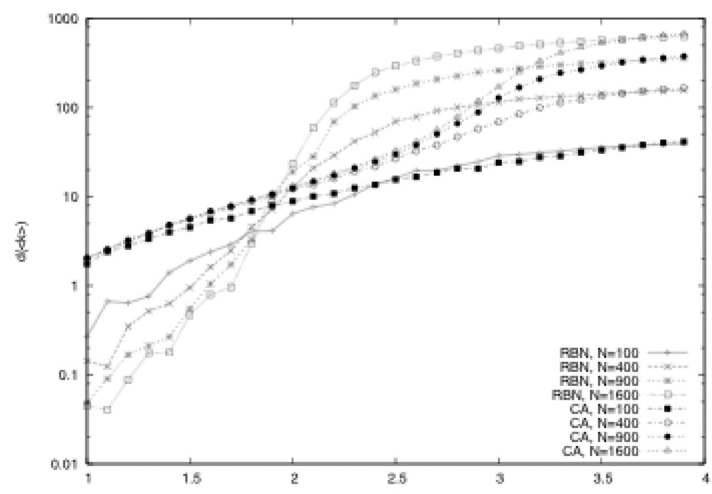
Figure 5. Average Hamming distance (damage) $\langle d \rangle$ *after 200 system updates, averaged over 100 randomly generated networks for each value of* $\langle K \rangle$*, with 100 different random initial conditions and a damage size of 10 nodes for each network. See text for discussion*



DISCUSSION

We hypothesize that the particular behavior can be explained by the percolation limit of the cellular automata. Da Silva et al. (Silva, Hanson, & Roux, 1989) found that the link probability at the percolation limit is approximately $p \sim 0.6$, which means that the average connectivity at the perco-

Figure 6. Average Hamming distance (damage) $\langle d \rangle$ *after 200 system updates, averaged over 100 randomly generated networks for each value of* $\langle K \rangle$*, with 100 different random initial conditions and a damage size of 20 nodes for each network. See text for discussion*



lation limit in our CA topology with a maximum of 4 neighbors is given by $\langle k \rangle$=4*p*=2.4. This value corresponds to the experimentally observed value where the damage spreading suddenly becomes dependent of the system size. Because of the local CA connectivity, there are lots of disconnected components below the percolation limit. Below this limit, the damage spreading is thus very slow and limited by the disconnected components, reason why it is essentially independent of system size. Above the percolation limit, the CA suddenly becomes connected and damage spreading becomes therefore dependent on the system size. For larger damage events, such as 10 or 20, damage becomes more dependent on system size even below the percolation limit because there is a higher probability that damage is induced in several disconnected components at the same time.

More recently, Lu & Teuscher (Lu & Teuscher, 2009) have also investigated the damage spreading in spatial random networks with extreme local connections and with small-world topologies. They found that spatially local connections change the scaling of the relevant component at very low connectivities ($K \ll 1$), that there is no critical connectivity of stability K_s for locally connected networks, and that the critical connectivity K_s changes to lower values compared to random networks for small-world networks.

In summary: for single-node damage events, CAs offer system-size independent damage spreading for up to about $\langle K \rangle = 2.4$ (which corresponds to the percolation limit), however, this particular behavior disappears for larger damage events. We conclude that in the general case, CAs do not possess a characteristic connectivity K_s, where damage spreading is independent of the system size N. Such a connectivity, however, exists for both RBNs and RTNs, which makes them particularly suitable as a computing model in an environment with high error probabilities or systems with low system component reliabilities. Examples are logical gates based on bio-molecular components (Benenson et al., 2001), where high failure rates can be expected.

COMPLEX NETWORKS AND WIRING COSTS

Most real networks, such as brain networks (Sporns, Chialvo, Kaiser, & Hilgtag, 2004; Eguéluz, Chialvo, Cecchi, Baliki, & Apkarian, 2005), electronic circuits (Cancho, Janssen, & Sole, 2001), the Internet, and social networks share the so-called *small-world* (SW) property (Watts & Strogatz, 1998). Compared to purely locally and regularly interconnected networks (such as for example the CA interconnect of Figure 2), small-world networks have a very short average distance (measured as the number of edges to traverse) between any pair of nodes, which makes them particularly interesting for efficient communication.

The classical Watts-Strogatz small-world network (Watts & Strogatz, 1998) is built from a regular lattice with only nearest neighbor connections. Every link is then rewired with a *rewiring probability p* to a randomly chosen node. Thus, by varying *p*, one can obtain a fully regular (*p*=0) and a fully random (*p*=1) network topology. The rewiring procedure establishes "shortcuts" in the network, which significantly lower the average distance (i.e., the number of edges to traverse) between any pair of nodes. In the original model, the length distribution of the shortcuts is uniform since a node is chosen randomly. If the rewiring of the connections is done proportional to a power law, $l^{-\alpha}$, where l is the wire length, then we obtain a *small-world power-law network*. The exponent α affects the network's communication characteristics (Kozma, Hastings, & Korniss, 2005) and navigability (Kleinberg, 2000a), which is better than in the uniformly generated small-world network. One can think of other distance-proportional distributions for the rewiring, such as for example a Gaussian distribution, which has been found between certain layers of the rat's neocortical pyramidal neurons (Hellwig, 2000).

In a real network, it is fair to assume that local connections have a lower cost (in terms of the associated wire-delay and the area required) than long-distance connections. Physically realizing small-world networks with uniformly distributed long-distance connections is thus not realistic and distance, i.e., the wiring cost, needs to be taken into account, a perspective that recently gained increasing attention in complex networks community (Petermann & Rios, 2006). On the other hand, a network's topology also directly affects how efficient problems can be solved.

Teuscher (Teuscher, 2007; Teuscher et al., 2009) has pragmatically and experimentally investigated important design trade-offs and properties of an irregular, abstract, yet physically plausible 3D small-world interconnect fabric that is inspired by modern network-on-chip paradigms. The results confirm that (1) computation in irregular assemblies is a promising and disruptive computing paradigm for self-assembled nano-scale electronics and (2) that 3D small-world interconnect fabrics with a power-law decaying distribution of shortcut lengths are physically plausible and have major advantages over local 2D and 3D regular topologies, such as CA interconnects. More recently, Teuscher et al. (Teuscher et al., 2009) proposed two simple growth models for non-classical nanowire interconnects with the goal to both the cost and the communication of such networks. They have also used an evolutionary algorithm to evolve optimal network topologies under given cost and communication constraints.

DISCUSSION

There is a trade-off between (1) the physical realizability and (2) the communication characteristics for a network topology. A locally and regularly interconnected topology, such as that of a CA, is in general easy to build (especially for top-down engineered CMOS technology) and only involves minimal wire and area cost (as for example shown by Zhirnov et al. (V. Zhirnov et al., 2008)), but it offers poor global communication characteristics and scales-up poorly with system size. On the other hand, a random topology, such as that of RBNs or RTNs, scales-up well and has a very short-average path length, but it is not physically plausible because it involves costly long-distance connections established independently of the Euclidean distance between the nodes. The RBN and RTN topologies we consider here as thus extremes, such as CA topologies. The ideal topology lies in between: small-world topologies with a distance-dependent distribution of the connectivity. Such topologies are located in a unique spot in the design space and also offer two other highly relevant properties (Kleinberg, 2000b; Teuscher, 2007): (1) efficient navigability and thus potentially efficient routing, and (2) robustness against random link removals. For these reasons, we can conclude that small-world graphs are the most promising interconnects for future massive scale devices. This outcome is further supported by the quantitative analysis presented in (Teuscher at al., 2009): unstructured network-on-chip topologies obtained by the growth models show specific wire-length distributions that are beneficial for the communication and minimize the wiring cost.

These outcomes seem contradictory with Bilardi & Preparata (Bilardi & Preparata, 1995) argument that future architectures will naturally be driven toward regular and locally interconnected architectures. However, our findings have a different starting point, namely a set of computing nodes randomly distributed in space, which is motivated by the way we imagine such a system would be fabricated. Given that physical arrangement, we then ask the question of what the optimal interconnect is.

SOLVING TASKS WITH RANDOM BOOLEAN NETWORKS

In their pioneering work, Carnevali & Patarnello (Carnevali & Patarnello, 1989) used two global stochastic optimization techniques, *Simulated Annealing (SA)* and *Genetic Algorithms (GA)*, to train feedforward random Boolean networks to solve computational tasks. Paternello and Carnevali showed that global stochastic optimization is able to train networks for simple memorization and generalization tasks. Teuscher et al. (Teuscher et al., 2007) used the same techniques to train random Boolean networks and showed that they can successfully solve simple computational tasks as well. Their preliminary results also hinted at the effect of the critical connectivity on the training and generalization performance. In this section we first replicate the above-mentioned results and then analyze the dynamics of learning and generalization in RBNs. In particular, we investigate how the network connectivity influences these dynamics.

EXPERIMENTAL SETUP

For our purpose, we use genetic algorithms to train the networks for solving the bitwise AND, the even-odd, and the mapping task. The bitwise AND task consists of calculating the bitwise logical AND of two l-bit binary numbers. The networks thus have a $2 \times l$-bit inputs and an l-bit output. For the even-odd task, we have an l-bit binary input and a one-bit output. The output is one if the number of 1s in the input is an odd number, and 0 otherwise. The mapping task consists of an l-bit input and an l-bit output. The task consists of producing the same number of 1s in the output as there are in the input, but without considering the order of the bits. In order to perform these tasks with RBNs, we add additional input and output nodes to the regular network nodes. These nodes are connected randomly to other nodes in the network. While the input nodes only provide the input signals to the network, the output nodes may also compute a Boolean function.

In order to use GAs, we encode the network into a bit-stream that consists of both the network's adjacency matrix and the Boolean transfer functions for each node. The genetic operators consist of the standard mutation and one-point crossover operators that are applied to the genotypes in the network population. We further define a fitness function f and a generalization function g. For an input space N of size n and an input sample M of size m we write:

$$E_M = \frac{1}{m}\sum_{j\in M} d(j),\ f = 1 - E_M$$

where $d(j)$ is the Hamming distance between the network output for the j-th input in the random sample from the input space and the expected network output for that input. Similarly, we write:

$$E_N = \frac{1}{n}\sum_{i\in n} d(i),\ g = 1 - E_N$$

where $d(i)$ is the Hamming distance between the network output for the i-th input from the entire input space and the expected network output for that input.

The simple genetic algorithm we use to train the network is as following:

1. Create a random initial population of S networks.
2. Evaluate the performance of the networks on a random sample of the input space.
3. Apply the genetic operators to obtain a new population.
4. Continue with steps 2 and 3 until at least one of the networks achieves a perfect fitness or after G_{max} generations are reached.

While optimizing feedforward networks, we have to make sure that the mutation and crossover operators do not violate the feedforward topology of the network. Also, there is a subtle difference between feedforward RBNs and general (recurrent) RBNs: since recurrent RBNs can have feedback loops and thus have memory, we have to take into account the dynamical properties of the network and run them long enough to ensure they fall into an attractor. The network size N and the connectivity K determine the length of the transients, however, due to ambiguous outputs on cyclic attractors, we further calculate the majority activation of the output nodes over a number of time steps equal to the size of the network.

Carnevali & Patarnello (Carnevali & Patarnello, 1989) introduced the notion of *learning probability* as a way of describing the learning and generalization capability of their feedforward random networks. They defined the learning probability as the probability of the training process yielding a network with perfect generalization, given that the training achieves perfect fitness on a sample of the input space. To calculate this measure, we run the training process r times and store both the f and g values.

RESULTS

Figure 7 and 8 show the learning probability of feedforward RBNs on the even-odd and the bitwise AND task for $K = 2$ networks. As one can see, as the size of the input space increases, the training process requires a smaller number of training examples to achieve a perfect learning probability. We tune the GA parameters, such as the variation rate and the maximum number of generations

Figure 7. The learning probability of feedforward networks on the bitwise AND task on different problem sizes. As the number of input bits increases, the learning process requires a smaller fraction of the input space during the training to achieve a perfect learning score. $N = 50$, $K = 2.0$, $G_{max} = 3000$, initial population size = 50, crossover rate = 0.6, mutation rate = 0.3. The GA was repeated for 700 times

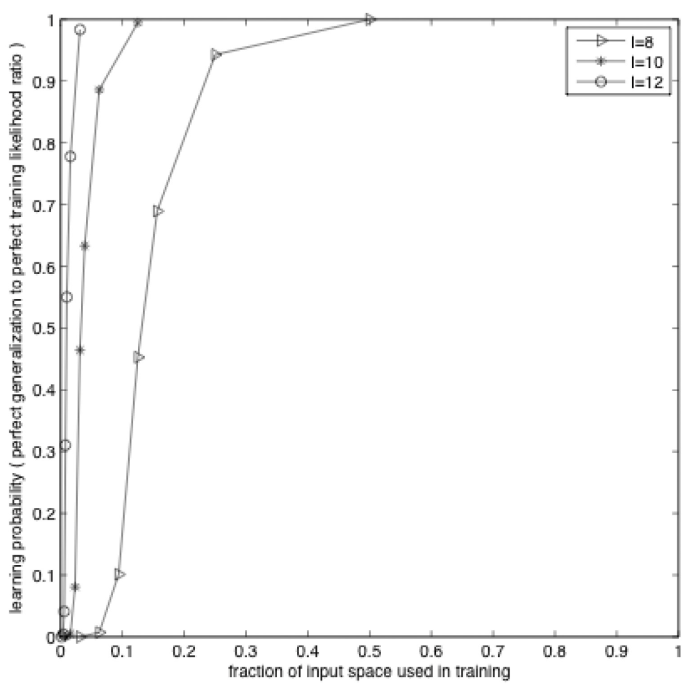
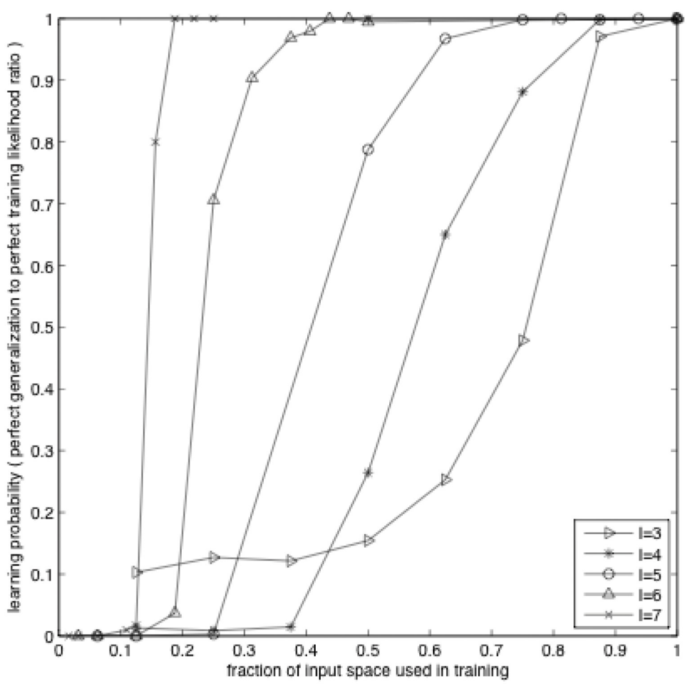


Figure 8. The learning probability of feedforward networks on the even-odd task for various input sizes. As the number of input bits increases, the learning process requires a smaller fraction of the input space during the training to achieve a perfect learning score. For I = 3, some of the networks can solve a significant number of patters without training because the task is too easy. N = 50, K = 2.0, G_{max} = 3000, initial population size = 50, crossover rate = 0.6, mutation rate = 0.3. The GA was repeated for 700 times

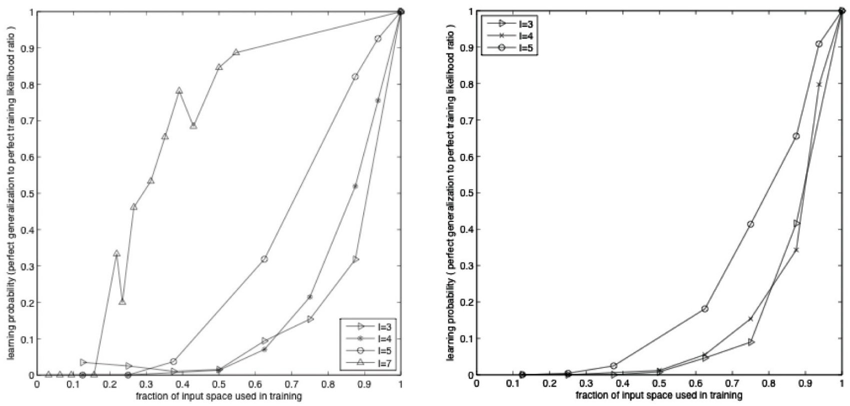


experimentally, depending on how quickly we achieve perfect fitness on average. These result directly confirm Carnevali & Patarnello (Carnevali & Patarnello, 1989) experiments.

Next, we create the same learning probability curves for RBNs to show that they too can solve these simple tasks. Figure 9 illustrates the results for the even-odd task. Similar to feedforward networks, we observe that as the problem size increases, a lower proportion of training samples is required to achieve high learning probabilities. On the other hand, we observe that it is more difficult to get the network to learn the tasks. The network reaches a learning probability of 1 only when we use the entire input space during the training phase. The average node inputs $\langle K \rangle$ in this experiment was 2.0, so all networks were in the critical regime. To see how the three connectivity regimes (i.e., subcritical, critical, supercritical) affect the learning probability, we generated network populations of various $\langle K \rangle$ to solve the even-odd task with five input bits.

Figure 10 (left) shows the learning probability of RBNs with different average *K*s for the even-odd task for five inputs (i.e., the input space has 32 patterns). As one can see, networks with supercritical connectivity ($\langle K \rangle = 3.0$) achieve a lower learning probability than the networks at the "edge of chaos" ($\langle K \rangle = 2.0$ and $\langle K \rangle = 2.5$) and at subcritical connectivities ($\langle K \rangle = 1.0, 1.5, 1.875, 1.95$). However, networks with subcritical connectivity show a interesting jump from zero to perfect learning probability when the training sample size reaches 0.6. For $\langle K \rangle = 2.0, 2.5$, and 3.0, the network

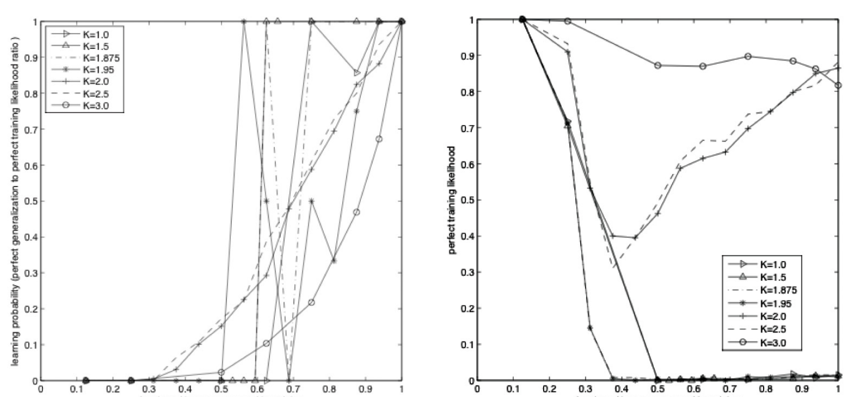
Figure 9. (left) The learning probability of RBNs for the even-odd task for various input sizes. N = 20, <K> = 2, G_{max} = 500. To calculate the likelihood we repeated the GA 500 times. (right) The learning probability of RBNs for the mapping task for various input sizes. N = 40, <K> = 2, G_{max} = 1000. We calculated the likelihoods over 260 runs. For both the mapping and even-odd tasks we used an initial population size of 50, a crossover rate of 0, and mutation rate of 0.7 for the GA



achieves some level of learning probability with smaller sample sizes (Figure 10 (left)).

Figure 10 (right) shows the training probability for the even-odd task. We see that subcritical K networks make the training very difficult. As the input sample size increases, the amount of information in the samples also does and subcritical networks do not have the flexibility to learn the task easily. On the other hand, networks in the complex and chaotic regime can easily learn the task. Moreover, although the critical $\langle K \rangle$ = *2* networks have more difficulty in learning the samples, they can generalize better, and thus have a higher learning probability than the networks with supercritical connectivity, which simply memorize the patterns they have learned.

Figure 10. (left) Learning probability of RBNs with different average K for even-odd task with five inputs. (right) Training probability of RBNs with different K for even-odd task with five inputs. N = 20, G_{max} = 500, initial population size = 50, crossover rate = 0, mutation rate = 0.7. We calculated the likelihood over 400 runs



DISCUSSION

We have shown that both feedforward and general (recurrent) RBNs can solve simple tasks and are able to generalize. While critical networks generalize better then subcritical and supercritical networks, the solution is harder to find. Our results show that there is great potential in training such networks more complex tasks and that the control of statistical properties of the network topology during the learning phase is an important parameter of generalization success.

MANUFACTURING ISSUES

As Chen et al. (Chen et al., 2003) state, "[i]n order to realize functional nano-electronic circuits, researchers need to solve three problems: invent a nanoscale device that switches an electric current on and off; build a nanoscale circuit that controllably links very large numbers of these devices with each other and with external systems in order to perform memory and/or logic functions; and design an architecture that allows the circuits to communicate with other systems and operate independently on their lower-level details."

While we can currently build switching devices in various technologies besides CMOS (see (V. V. Zhirnov, Hutchby, Bourianoff, & Brewer, 2006; Bourianoff, Brewer, Cavin, Hutchby, & Zhirnov, 2008; Hutchby, Cavin, Zhirnov, Brewer, & Bourianoff, 2008) for an overview), one of the remaining challenges is to assemble and interconnect these switching devices (or logic functions) to larger systems, and ultimately to design a computing architecture that allows to perform reliable computations. As mentioned before, there is little consensus in the research community on what type of technology and computing architecture holds most promises for the future.

The motivation for investigating randomly assembled interconnects and computing architectures can be summarized by the following observations:

- long-range and global connections are costly (in terms of wire delay and of the chip area used) and limit system performance (Ho et al., 2001);
- it is unclear whether a precisely regular and homogeneous arrangement of components is needed and possible on a multi-billion-component or even Avogadro-scale assembly of nano-scale components (Tour et al., 2002)
- "[s]elf-assembly makes it relatively easy to form a random array of wires with randomly attached switches" (V. V. Zhirnov & Herr, 2001); and
- building a perfect system is very hard and expensive

We have hypothesized in (Teuscher & Hansson, 2008) and (Teuscher, 2007) that bottom-up self-assembled electronics based on conductive nanowires or nanotubes can lead to the random interconnect topologies we are interested in, however, several questions remain open and are part of a 3-year interdisciplinary research project at Los Alamos National Laboratory (LANL). Our approach consists in using a hybrid assembly (as others explore as well, e.g., (Ferry, 2008)), where the functional building blocks will still be traditional silicon in a first step, while the interconnect is made up from self-assembled nanowires. Nanowires can be grown in various ways using diverse materials, such as metals and semiconductors. We have chosen a novel way to grow conductive nanowires, which Wang et al. (Wang, Li, Akhadov, & Jia, 2007) at LANL have pioneered and demonstrated: Ag nanowires can be fabricated on top of conducting polyaniline polymer membranes via a spontaneous electrodeless deposition (self-assembly) method. We hypothesize that this will allow to densely interconnect silicon components in a simple and cheap way with specific distance-dependent wire-length distributions. We believe that this approach will ultimately allow us to easily and cheaply fabricate RBN-like computing architectures.

Random threshold networks, on the other hand, could be rather straightforwardly and efficiently implemented with resonant tunneling diode (RTD) logic circuits (see e.g., (Pettenghi, Avedillo, & Quintana, 2008)), and represent a very interesting alternative to conventional Boolean logic gates. The reported results in this paper on random threshold networks can thus directly be applied to the implementation of such devices. There has been a significant body of research in the area of threshold logic in the past (see e.g., (Muroga, 1971)), but to the best of our knowledge, random threshold networks have not been considered as computing models for future and emerging computing machines.

CONCLUSION

The central claim of this paper is that locally interconnected computing architectures, such as cellular automata (CA), are in general not appropriate models for large-scale and general-purpose computations. We have supported this claim with recent theoretical results on the complex dynamical behavior of discrete random dynamical networks, their robustness to damage events as the system scales up, their ability to efficiently solve tasks, and their improved transport characteristics due to the short average path length. In a nutshell, the arguments why we believe that CAs are *not* promising architectures for future information-processing devices, are as following:

- their local interconnect topology is not small-world and has thus worse global transport characteristics (than small-world or random graphs), which directly affects the effectiveness of how general-purpose algorithmic tasks can be solved.
- in terms of a complex dynamical system, they operate in the supercritical regime () with the widely used von Neumann neighborhood, which makes them sensitive to initial conditions;
- they do not generally have a characteristic connectivity K_s, where damage spreading is independent of system size, which makes a system inherently robust; and
- it is unclear whether a precisely regular and homogeneous arrangement of components is possible at the scale of future information processing devices.

We have assessed RBNs and RTNs as alternative models, however, as we have seen above, they come at a serious cost: the uniform probability to establish connections with any node in the system independent of the Euclidean distance between them is not physically plausible and too expensive in terms of wiring cost. The ultimate interconnect topology is small-world and has a distance-dependent distribution of the wires (Teuscher, 2007; Teuscher & Hansson, 2008; Petermann & Rios, 2006). It was also been confirmed recently (Lu & Teuscher, 2009) that a critical connectivity K_s exists for such small-world networks as well.

OPEN QUESTIONS AND UNADDRESSED ISSUES

Naturally, there are a number of open questions and issues that we have not addressed because they are beyond the scope of this paper. In particular, an irregular topology with random logical functions makes the mapping of a given digital circuit much harder, if not impossible in certain cases. On the other hand, a regular interconnect topology clearly makes the mapping task easier. We believe, however, that automated design tools can address this challenge. After all, computation in random assemblies is not completely new and has been more or less successfully tried by others, e.g., (Patwardhan, Dwyer, Lebeck, & Sorin, 2006; Lawson & Wolpert, 2006, Tour et al., 2002), however in different contexts and with a different perspective in mind than we have presented here.

We have deliberately *not* focused on any particular application in this paper because our results are independent of the application. However, it is noteworthy that locally interconnected CAs have been proven to outperform other general purpose architecture on very specific applications. A good example are cellular neural networks (CNNs) (Chua & Roska, 2002), which, e.g., allow to perform certain image processing tasks orders of magnitude faster than any other machine.

Further, it is unknown at this point how exactly our findings fit into the interconnect predictions made by Rent's rule, however, the rule may not be applicable to our non-traditional circuits since it is based on empirical results. Further research on this is planned.

Last but not least, we would like to mention that, although we have only considered 2D arrangements and interconnects here for simplicity, the future is clearly 3D (e.g., see (Pavlidis & Friedman, 2007)). The main reason is that the average wire length in 3D is shorter than in 2D interconnects.

OUTLOOK

We believe that computation in random self-assemblies of simple components and interconnections is a highly appealing paradigm, both from the perspective of fabrication as well as performance and robustness. Future work will focus on (1) the manufacturing issues, (2) appropriate design methodologies, (3) addressing the mapping issues, and (4) more realistic models, which will allow to better assess the performance and cost, and (5) specific applications.

REFERENCES

Albert, R., & Barabási, A. L. (2000). Dynamics of complex systems: Scaling laws for the period of Boolean networks. *Physical Review Letters*, *54*, 5660–5663. doi:10.1103/PhysRevLett.84.5660

Avouris, P., Chen, Z., & Perebeinos, V. (2007). Carbon-based electronics. *Nature Nanotechnology*, *2*, 605–615. doi:10.1038/nnano.2007.300

Bastolla, U., & Parisi, G. (1998). Relevant elements, magnetization and dynamical properties in Kauffman networks: A numerical study. *Physica D. Nonlinear Phenomena*, *115*(3–4), 203–218. doi:10.1016/S0167-2789(97)00243-1

Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Leivneh, Z., & Shapiro, E. (2001). Programmable and autonomous computing machine made of biomolecules. *Nature*, *414*(6862), 430–434. doi:10.1038/35106533

Bilardi, G., G., & Preparata, F. P. (1995). Horizons of Parallel Computation. *Journal of Parallel and Distributed Computing*, *27*(2), 172–182. doi:10.1006/jpdc.1995.1080

Bourianoff, G., Brewer, J. E., Cavin, R., Hutchby, J. A., & Zhirnov, V. (2008). Boolean logic and alternative information-processing devices. *IEEE Computer*, *41*(5), 38–46.

Cancho, R. F. i, Janssen, C., & Sole, R. V. (2001). Topology of technology graphs: Small world patterns in electronic circuits. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *64*, 046119. doi:10.1103/PhysRevE.64.046119

Carnevali, S., & Patarnello, P. (1989). Learning capability of Boolean network In Aleksander, E. (Ed.), *Neural computing architecture: the design of brain-like machines* (pp. 117–129). Cambridge, MA: The MIT Press.

Chen, Y., Jung, G. Y., Ohlberg, D. A. A., Li, X., Steward, D. R., & Jepperson, J. O. (2003). Nanoscale molecular-switch crossbar circuits. *Nanotechnology*, *14*, 462–468. doi:10.1088/0957-4484/14/4/311

Chua, L. O., & Roska, T. (2002). *Cellular neural networks & visual computing*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511754494

da Silva, L. R., Hanson, A., & Roux, S. (1989). Spreading of damage in deterministic cellular automata as a percolation problem. *Europhysics Letters*, *8*(1), 47–52. doi:10.1209/0295-5075/8/1/009

Davis, J. A., Venkatesan, R., Kaloyeros, A., Beylansky, M., Souri, S. J., & Banerjee, K. (2001). Interconnect limits on gigascale integration (GSI) in the century. *Proceedings of the IEEE*, *89*(3), 305–324. doi:10.1109/5.915376

Derrida, B., & Pomeau, Y. (1986). Random networks of automata: A simple annealed approximation. *Europhysics Letters*, *1*(2), 45–49. doi:10.1209/0295-5075/1/2/001

Eguéluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., & Apkarian, A. V. (2005). Scale-free brain functional networks. *Physical Review Letters*, *94*, 018102. doi:10.1103/PhysRevLett.94.018102

Ferry, D. K. (2008). Nanowires in nanoelectronics. *Science*, *319*(5863), 579–580. doi:10.1126/science.1154446

Gershenson, C. (2003). Classification of random Boolean networks. In R. K. Standish, M. A. Bedau, & H. A. Abbass (Eds.), *Artificial life VIII. Proceedings of the eight international conference on artificial life* (pp. 1–8). Cambridge, MA: A Bradford Book, MIT Press.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biological Cybernetics*, *82*, 111–121. doi:10.1007/PL00007964

Ho, R., Mai, K. W., & Horowitz, M. A. (2001). The future of wires. *Proceedings of the IEEE*, *89*(4), 490–504. doi:10.1109/5.920580

Hutchby, J. A., Cavin, R., Zhirnov, V., Brewer, J. E., & Bourianoff, G. (2008). Emerging nanoscale memory and logic devices: A critical assessment. *IEEE Computer*, *41*(5), 28–32.

Kauffman, S. A. (1968). Metabolic stability and epigenesis in randomly connected genetic nets. *Journal of Theoretical Biology*, *22*, 437–467. doi:10.1016/0022-5193(69)90015-0

Kauffman, S. A. (1984, January). Emergent properties in random complex automata. *Physica D. Nonlinear Phenomena*, *10*(1–2), 145–156. doi:10.1016/0167-2789(84)90257-4

Kauffman, S. A. (1993). *The origins of order: Self–organization and selection in evolution*. New York, Oxford: Oxford University Press.

Kleinberg, J. K. (2000a). Navigation in a small world. *Nature*, *406*, 845. doi:10.1038/35022643

Kleinberg, J. K. (2000b). The small-world phenomenon: An algorithmic perspective. In *Proceedings of the annual acm symposium on theory of computing* (pp. 163–170).

Kozma, B., Hastings, M. B., & Korniss, G. (2005). Diffusion processes on power-law small-world networks. *Physical Review Letters*, *95*, 018701. doi:10.1103/PhysRevLett.95.018701

Lawson, J., & Wolpert, D. H. (2006). Adaptive programming of unconventional nano-architectures. *Journal of Computational and Theoretical Nanoscience*, *3*, 272–279.

Lu, Q., & Teuscher, C. (2009). Damage Spreading in Spatial and Small-world Random Boolean Networks, 2009. In revision. arxiv:cond-mat/0904.4052

Meindl, J. D. (2003). Interconnect opportunities for gigascale integration. *IEEE Micro*, *23*(3), 28–35. doi:10.1109/MM.2003.1209464

Muroga, S. (1971). *Threshold logic and its applications*. New York: Wiley-Interscience.

Patwardhan, J. P., Dwyer, C., Lebeck, A. R., & Sorin, D. J. (2006). NANA: A nano-scale active network architecture. *Journal on Emerging Technologies in Computing Systems*, *2*(1), 1–30. doi:10.1145/1126257.1126258

Pavlidis, V. F., & Friedman, E. G. (2007). 3-D topologies for networks-on-chip. *IEEE Transactions of Very Large Scale Integration (VLSI) Syste,s, 15*(10), 1081–1090.

Petermann, T., & Rios, P. D. L. (2006). Physical realizability of small-world networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *73*, 026114. doi:10.1103/PhysRevE.73.026114

Pettenghi, M., Avedillo, M. J., & Quintana, J. M. (2008). Using multi-threshold threshold gates in RTD-based logic design: A case study. *Microelectronics Journal*, *39*(2), 241–247.

Rohlf, T., & Bornholdt, S. (2002). Criticality in random threshold networks: annealed approximation and beyond. *Physica A*, *310*, 245–259. doi:10.1016/S0378-4371(02)00798-7

Rohlf, T., Gulbahce, N., & Teuscher, C. (2007). Damage spreading and criticality in finite random dynamical networks. *Physical Review Letters*, *99*(24), 248701. doi:10.1103/PhysRevLett.99.248701

Samuelsson, B., & Troein, C. (2003). Superpolynomial growth in the number of attractors in Kauffman networks. *Physical Review Letters*, *90*(098701).

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgtag, C. C. (2004). Organization, development, and function of complex brain networks. *Trends in Cognitive Sciences*, *8*(9), 418–425. doi:10.1016/j.tics.2004.07.008

Teuscher, C. (2007). Nature-inspired interconnects for emerging large-scale network-on-chip designs. *Chaos (Woodbury, N.Y.)*, *17*(2), 026106. doi:10.1063/1.2740566

Teuscher, C., Gulbahce, N., & Rohlf, T. (2007, Jan 3–6). Learning and generalization in random Boolean networks. In *Dynamic days 2007: International conference on chaos and nonlinear dynamics.* Boston, MA.

Teuscher, C., Gulbahce, N., & Rohlf, T. (2009b). An assessment of random dynamical network automata for nanoelectronics. *International Journal of Nanotechnology and Molecular Computation*, *1*(4), 39–57.

Teuscher, C., & Hansson, A. A. (2008). Non-traditional irregular interconnects for massive scale SoC. In *Proceedings of the IEEE Iinternational Symposium on Circuits and Systems (ISCAS)* (pp. 2785–2788).

Teuscher, C., Parashar, N., Mote, M., Hergert, N., & Aherne, J. (2009). Wire Cost and Comunication Analysis of Self-Assembled Interconnect Models for Networks-on-Chip. In *Proceedings of the Second International Workshop on Network on Chip Architectures* (NoCArc 2009), 83–88.

Tour, J., Zandt, W. L. V., Husband, C. P., Husband, S. M., Wilson, L. S., & Franzon, P. D. (2002). Nanocell logic gates for molecular computing. *IEEE Transactions on Nanotechnology*, *1*(2), 100–109. doi:10.1109/TNANO.2002.804744

von Neumann, J. (1966). *Theory of self-reproducing automata*. Urbana, IL: University of Illinois Press.

Wang, H. L., Li, W., Akhadov, E., & Jia, Q. (2007). Electrodeless deposition of metals and metal nanoparticle using conducting polymers. *Chemistry of Materials*, *19*, 520. doi:10.1021/cm0619508

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442. doi:10.1038/30918

Wolfram, S. (1984, December 4). Cellular automata as models of complexity. *Nature*, *311*, 419–424. doi:10.1038/311419a0

Zhirnov, V., Cavin, R., Leeming, G., & Galatsis, K. (2008). An assessment of integrated digital cellular automata architectures. *IEEE Computer*, *41*, 38–43.

Zhirnov, V. V., & Herr, D. J. C. (2001, January). New frontiers: Self-assembly in nanoelectronics. *IEEE Computer*, *34*, 34–43.

Zhirnov, V. V., Hutchby, J. A., Bourianoff, G. I., & Brewer, J. E. (2006). Emerging research logic devices. *IEEE Circuits & Devices Magazine*, *21*, 36–46.

ENDNOTE

1 This paper is a significant extension of work first presented at the Nanoarch 2008 workshop, Anaheim, CA, USA, Jun 12-13, 2008, and subsequently published in (Teuscher et al., 2009b). This present chapter contains a new section on "Solving Tasks with Random Boolean Networks."

This work was previously published in Theoretical and Technological Advancements in Nanotechnology and Molecular Computation: Interdisciplinary Gains, edited by Bruce MacLennan, pp. 295-314, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 91
Creative Regions in Europe:
Exploring Creative Industry Agglomeration and the Wealth of European Regions

Blanca de-Miguel-Molina
Universitat Politècnica de València, Spain

José-Luis Hervás-Oliver
Universitat Politècnica de València, Spain

Rafael Boix
Universitat de València, Spain

María de-Miguel-Molina
Universitat Politècnica de València, Spain

ABSTRACT

This chapter examines the existence of regional agglomerations of manufacturing, service and creative industries, and the relationship between these industries and the wealth of regions. Through an analysis of 250 European regions, three important conclusions can be inferred from the results obtained in this chapter. The first is that creative industries play an important role in the wealth of the regions. The second is that the most creative regions are characterized by having more high-tech manufacturing industries than the rest of the regions, although the number of low-tech manufacturing firms is similar. Lastly, in the richest regions, a greater share of high-tech manufacturing industries coexists with a greater share of creative industries. The importance of this chapter resides in the fact that up until now no analysis has demonstrated that creative industries are the most important industries in regional wealth.

1. INTRODUCTION

The localization of creative industries, as highlighted by the work of Stam et al. (2008), Cooke (2008), Lazzeretti et al. (2008), Capone (2008) and Power and Nielsen (2010), is an area of increasing importance in the literature on geographic agglomerations. These industries are in fact groupings of specific sectors of low-technology manufacturing and knowledge-intensive services, which is why their importance is related to the ever-increasing dependence of manufacturing sectors on service industries (Peneder et al., 2003; Pilat and Wölfl,

DOI: 10.4018/978-1-4666-1945-6.ch091

2005; Drejer and Vinding, 2005; Wood, 2006; Aslesen and Isaksen, 2007b), and on what we could call the *Knowledge and Service Economy* (Windrum and Tomlinson, 1999; Bishop, 2008; Aslesen and Isaksen, 2007a; Aslesen and Isaksen 2007b; Strambach 2008).

Existing studies show which activities can be included in creative industries (UNCTAD, 2010), and why these industries form agglomerations (Lazzeretti et al., 2009; Lorenzen and Frederiksen, 2008). However, although the existence of relationships between manufacturing and services agglomerations in their different definitions of high and low (Leydesdorff et al., 2006; Leydesdorff and Fritsch, 2006; Vence-Deza and González-López, 2008; Heidenreich, 2009) have been the aim of some analyses, the above mentioned analyses have not been carried out to examine the possible relationship between manufacturing agglomerations – including both high and low-tech industries – and creative industries. This paper attempts to fill this gap. In addition, the core importance of the paper is based on the fact that it points out empirically how important creative industries are in developing economies and bringing prosperity to European regions. The results obtained have implications for academia as well as for policymakers.

The empirical study is based on a sample of 250 regions in 24 European countries. The data was taken from Eurostat's *Structural Business Statistics* and *Regional Economic Accounts* databases. The data was used to evaluate manufacturing, service and creative industry agglomerations based on the *Location Quotient* (LQ) of firms by regions. Following this, the relationship between these agglomerations and GDP was verified.

Three important conclusions are pointed out. The first is that creative industries play an important role in the wealth of a region. The second is that the most creative regions are characterized by having more high-tech manufacturing sectors than others, although they have a similar share of low-tech manufacturing firms. The third is that in the richest regions a greater share of high-tech manufacturing industries coexists with a greater share of creative industries, thus showing the importance of creativity.

The outline that we have used in this paper is as follows: in Sections 2 we briefly summarize the recent basic theory on the study of maps of manufacturing, service and creative industry agglomerations to determine the localization patterns and the relationships they have on the aforementioned studies. In Section 3, we include the empirical study where we set out the variables used, the sources the data was extracted from and the methodology used for their study as well as the results obtained. Conclusions can be found in Section 4.

2. MAPS OF AGGLOMERATIONS IN MANUFACTURING, SERVICES AND CREATIVE INDUSTRIES

The maps of agglomerations in manufacturing, services and creative industries are representations of sectors which are located in a geographic zone, whether it is a city, region or country. Examples of these maps can be found in studies on the different terminologies used to name these agglomerations according to their characteristics: industrial districts, clusters, milieux and Local Production Systems. Therefore, we have maps of sectoral concentrations for most of the European countries (Becattini and Coltorti, 2006; Crouch and Farrel, 2001; Pitelis and Pseiridis, 2006; Boix and Trullén, 2011).

Recently, the study of maps of clusters and districts has acquired a new dimension, in which instead of limiting the localization of the different industries and services at separate levels, the maps are more aggregated. Examples of the former can be found in the studies carried out by Becattini and Coltorti (2006), who point out the different industrial districts in Italy – the shoe industry located in Brenta, Fermano Maceratese and Verona, the tile industry in Sassuolo and the textile industry in Prato and Biella, among others. In Spain, these types of studies have also been carried out by Boix and Trullén (2011), among others and the same has occurred in the majority of European countries and on other continents. At first, the agglomerations studied corresponded to industry while the service sector (O'Donoghue and Gleave, 2004) was incorporated later as its influence on the economies increased. Pilat and Wölfl (2005) have pointed out that this increase is the result of the relationship between manufacturing and service industries and because the former tends to subcontract some activities to firms with specialized services located in the same country or in a foreign country. Studies such as the one by Heidenreich (2009) found that regions specialize either in manufacturing or in services, while Wood (2006) pointed out the dependence that manufacturing has on services, and especially knowledge-intensive services (KIS).

In Table 1 we have included the sectors that come under manufacturing and services industries from the NACE Rev.2 classification and have separated them according to their technology and knowledge levels.

A specific case study of agglomerations which includes manufacturing and services of different types is that of *creative industries* (Stam et al. 2008; Lazzeretti et al., 2008; Capone, 2008; De Propis et al., 2009; Power and Nielsén, 2010). According to Pratt (2008), it was toward the end of the 1990s when this terminology began to be used in Europe, to be more precise, when the British Department for Culture, Media and Sport (DCMS) drew up its map of creative industries in 1998.

The question about which activities are creative was influenced by the inclusion of activities that involve copyright or other intellectual property (Towse, 2010). The most widely extended definition of creative industries is that of the DCMS (2009) which defined creative industries, as "those industries that are based on individual creativity, skill and talent. And which have the potential to create wealth and jobs through developing intellectual property." Howkins (2005) believes that it "is more consistent to restrict the term to an industry where intellectual labour is predominant and the result is intellectual property." UNCTAD (2010) considers creative industries as "any economic activity producing symbolic products with a heavy reliance on intellectual property and for as wide a market as possible." The difference is in the DCMS's use of the term "individual," which has been argued by some authors (Healy, 2002; Garnham, 2005; Pratt, 2005; O'Connor, 2007).

The most comprehensive taxonomy of creative industries, which is particularly appropriate to cross-country comparisons has been proposed by UNCTAD (2010). Table 2 contains the transformation to NACE Rev 2 of these activities, separating them by manufacturing or service industries. It has been verified that creative industries are made up of sectors pertaining to low-tech manufacturing and knowledge-intensive services, although the majority of the sectors included are service industries and especially KIS.

When comparing the definition of creative industries as per the British Department for Culture (Pratt 2008, DCMS 2009) with the characteristics attributed to KIS sectors (Nählinder, 2005; Doloreux et al., 2008; Strambach, 2008; Muller and Doloreux, 2009; Shearmur and Doloreux, 2009), both make reference to the talent and abilities of persons and firms to create knowledge (Larsen, 2001; Aslesen and Isaksen, 2007b).

Secondly, the most aggregated analysis studies high, medium and low-technology manufacturing sectors globally, sometimes separately (Cooke et al, 2007, chap. 8), and other times jointly (Robertson and Patel, 2007). The differences in technology levels are based on R&D intensity (Hatzichronoglou, 1997), which the OECD uses to establish its four types of manufacturing industries: a) high-technology, b) medium-high-technology, c) medium-low-technology and d) low-technology. What is taken into account in services is knowl-

Table 1. Aggregations of manufacturing and services based on NACE Rev 2, 2-digit level

Manufacturing		
High-technology		***21*** *Manufacture of basic pharmaceutical products and pharmaceutical preparations* ***26*** *Manufacture of computer, electronic and optical products*
Medium-high-technology		***20*** *Manufacture of chemicals and chemical products* ***27*** *Manufacture of electrical equipment* ***28*** *Manufacture of machineryand equipment n.e.c.* ***29*** *Manufacture of motor vehicles, trailers and semi-trailers* ***30*** *Manufacture of other transport equipment*
Medium-low-technology		***19*** *Manufacture of coke and refined petroleum products* ***22*** *Manufacture of rubber and plastic products* ***23*** *Manufacture of other non-metallic mineral products* ***24*** *Manufacture of basic metals* ***25*** *Manufacture of fabricated metal products, except machinery and equipment* ***33*** *Repair and installation of machinery and equipment*
Low-technology		***10*** *Manufacture of food products* ***11*** *Manufacture of beverages* ***12*** *Manufacture of tobacco products* ***13*** *Manufacture of textiles* ***14*** *Manufacture of wearing apparel* ***15*** *Manufacture of leather and related products* ***16*** *Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials* ***17*** *Manufacture of paper and paper products* ***18*** *Printing and reproduction of recorded media* ***31*** *Manufacture of furniture* ***32*** *Other manufacturing*
Services		
Knowledge-intensive services (KIS)	High-tech Knowledge-intensive services (HTKIS)	***59*** *Motion picture, video and television programme production, sound recording and music publishing activities* ***60*** *Programming and broadcasting activities* ***61*** *Telecommunications* ***62*** *Computer programming, consultancy and related activities* ***63*** *Information service activities* ***72*** *Scientific research and development*
	Other Knowledge-intensive services (OKIS) [(1)]	***50*** *Water transport* ***51*** *Air transport* ***58*** *Publishing activities* ***64*** *Financial service activities, except insurance and pension funding* ***65*** *Insurance, reinsurance and pension funding, except compulsory social security* ***66*** *Activities auxiliary to financial services and insurance activities* ***69*** *Legal and accounting activities* ***70*** *Activities of head offices; management consultancy activities* ***71*** *Architectural and engineering activities; technical testing and analysis* ***73*** *Advertising and market research* ***74*** *Other professional, scientific and technical activities* ***75*** *Veterinary activities* ***78*** *Employment activities* ***80*** *Security and investigation activities* ***84*** *Public administration and defence; compulsory social security* ***85*** *Education* ***86*** *Human health activities* ***87*** *Residential care activities* ***88*** *Social work activities without accommodation* ***90*** *Creative, arts and entertainment activities* ***91*** *Libraries, archives, museums and other cultural activities* ***92*** *Gambling and betting activities* ***93*** *Sports activities and amusement and recreation activities*

Table 1. Continued

Less-Knowledge-intensive services (LKIS)	***45*** *Wholesale and retail trade and repair of motor vehicles and motorcycles* ***46*** *Wholesale trade, except of motor vehicles and motorcycles* ***47*** *Retail trade, except of motor vehicles and motorcycles* ***49*** *Land transport and transport via pipelines* ***52*** *Warehousing and support activities for transportation* ***53*** *Postal and courier activities* ***55*** *Accommodation* ***56*** *Food and beverage service activities* ***68*** *Real estate activities* ***77*** *Rental and leasing activities* ***79*** *Travel agency, tour operator reservation service and related activities* ***81*** *Services to buildings and landscape activities* ***82*** *Office administrative, office support and other business support activities* ***94*** *Activities of membership organisations* ***95*** *Repair of computers and personal and household goods* ***96*** *Other personal service activities* ***97*** *Activities of households as employers of domestic personnel* ***98*** *Undifferentiated goods-and service-producing activities of private households for own use* ***99*** *Activities of extraterritorial organisations and bodies*

(1) KIS except High-tech knowledge-intensive services

Source: Eurostat: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/High-tech_statistics

Table 2. Aggregations of creative industries based on NACE Rev. 2. Adaptation to 2 digits.

Manufacturing	**Creative**	**Non-creative**
High-tech		21, 26
Medium-high tech		20, 27, 28, 29, 30
Medium-low tech		19, 22, 23, 24, 25, 33
Low-tech	14, 15, 18,	10, 11, 12, 13, 16, 17, 31, 32
Services	**Creative**	**Non-creative**
High-tech Knowledge-intensive services (HTKIS)	59, 60, 62, 72	61, 63
Other Knowledge-intensive services (OKIS)	58, 71, 73, 74, 90, 91, 93	50, 51, 64, 65, 66, 69, 70, 75, 78, 80, 84, 85, 86, 87, 88, 92
Less-Knowledge-intensive services (LKIS)		45, 46, 47, 49, 52, 53, 55, 56, 68, 77, 79, 81, 92, 94, 95, 96, 97, 98, 99

Source: UNCTAD (2010), DCMS (2009), KEA European Affairs (2006) and Lazzeretti et al. (2008)

edge, since the relationship existing between the manufacturing and service sectors allows the latter to transfer knowledge to the former, as well as to create it (Miles, 2008). This is why the analysis of knowledge-intensive services (Bishop, 2008) predominates in the services industry, since they are associated with the *knowledge-based economy* (Windrum and Tomlinson, 1999; Aslesen and Isaksen, 2007a; Bishop, 2008; Strambach, 2008).

Lastly, we find the analysis of the relationship between manufacturing and service agglomerations (Leydesdorff et al. 2006; Leydesdorff and Fritsch, 2006), to which the relationship between agglomerations and wealth – measured by GDP per inhabitant – is sometimes added (Heidenreich, 2009; Vence-Deza and González-López, 2008). However, the relationship with creative industries has not been sufficiently explored.

In the studies currently being carried out on maps at aggregate level, some patterns have been detected with respect to the importance of high-tech sectors in manufacturing and services in relation to the generation of knowledge and to regional wealth (measured by GDP per inhabitant).

There is not a common opinion (Leydesdorff and Fritsch, 2006; Leydesdorff et al., 2006; Robertson and Patel, 2007; Bishop, 2008; Vence-Deza and González-López, 2008; Heidenreich, 2009; The Center for Strategy and Competitiveness, 2009) about whether high-tech manufacturing sectors are the only relevant ones since their importance is usually associated with that of medium-high tech sectors, even though the presence of medium-tech industries is considered important for regional wealth. In the case of service sectors, however, the importance of knowledge-intensive services (KIS) and high-tech knowledge-intensive services (HTKIS, which are included in KIS) can indeed be observed.

Vence-Deza and González-López (2008) pointed out that the principal trend in European regions is towards geographic concentration of high-tech sectors in manufacturing and services and that this concentration takes place in the regions with the greatest GDP per inhabitant. However, Leydesdorff and Fritsch (2006) and Leydesdorff et al. (2006), in two studies for Germany and Holland, verified that medium-tech sectors are much more important than high-tech sectors in the knowledge base of a region or country. It has also been observed that high-tech knowledge-intensive services (HTKIS) are the most important for the service sector while other knowledge-intensive services (OKIS) have a lesser effect on the territorial knowledge base.

In Heidenreich's opinion (2009), the richest regions – those with a high GDP – are those which have a high percentage of jobs both in high-tech sectors as well as in medium high-tech manufacturing sectors and in knowledge-intensive services (KIS), but have a low percentage of jobs in low and medium-tech manufacturing sectors. The Center for Strategy and Competitiveness (2009) in its study on KIBS (knowledge-intensive business services) sectors in Europe found that regions with strong KIBS sectors had the highest prosperity levels in Europe. Contrary to this finding, Bishop (2008) argues that for the United Kingdom development depends on diversification rather than depending solely on KIS sectors or on any other single sector. It is not strange to reach this conclusion for the United Kingdom, which is one of the countries that places greatest importance on the role of creative industries. Therefore, the evidence is not really clear about what type of industries determine the wealth of regions.

Our next question refers to the relationship between industries and whether the studies existing at present show a regional trend towards specialization in high or low sectors.

Heidenreich (2009) found a pattern in European regions where low and medium low-tech (LMT) sectors coexist with high and medium-high sectors, but did not find a similar situation in manufacturing and services, i.e., in regions which specialize in manufacturing and services.

Accordingly, Robertson and Patel (2007) evidenced the compatibility between high-tech manufacturing and low-tech manufacturing industries. Therefore, in European countries, the presence of high-tech manufacturing industries is compatible with the presence of low-tech manufacturing industries.

Regarding to creative industries, Power (2002, 2003), Cooke (2008), Stam et al. (2008), Lazzeretti et al. (2008), Capone (2008), De Propris et al (2009), Baum et al. (2009), and Power and Nielsén (2010) have studied their location from an aggregated viewpoint. These studies conclude that creative industries tend to be located in the major urban areas of each country.-

In terms of the relationship between location and wealth, the DCMS creative industries concept expresses "their potential for wealth and job creation." Power and Nielsén (2010) found that creative and cultural industries are located in the wealthiest European regions, while Stam et al. (2008) showed that the presence of the creative class has a higher impact on employment growth than creative industries. Baum et al. (2009) also pointed out that locations need human capital if they intend to prosper in creative industries.

In our empirical study we have broadened these conclusions to include the study of creative regions which has not been carried out previously.

3. METHODOLOGY

a) Sample and Variables

The sample comprises 250 European regions. The data for this study was compiled from Eurostat's *Structural Business Statistics* (SBS) and *Regional Economic Accounts (REA)* databases and corresponds to 2008. The countries whose data was not available, such as Greece, Luxembourg and Malta, were not included.

The variables extracted were used to calculate the *Location Quotient* (LQ) for manufacturing (high, medium-high, medium-low and low tech), services (HTKIS, OKIS and LKIS) and creative industries for each region with relation to EU-27. The Location Quotient (LQ) is an indicator of the existence of industrial agglomerations in a region. The formula used in our study is presented in Exhibit 1.

Additionally, the data on GDP per inhabitant for each region was taken in order to examine its relationship with creative industries and to compare it with the results in reference to manufacturing and services.

The following statistical calculations were carried out based on the aforementioned limitations. First, the correlations between industries were calculated to find out in which cases there were positive or negative correlations as well as their correlation with respect to GDP per inhabitant. Second, a cluster analysis was carried out to see whether groups of regions by GDP per inhabitant and the number of creative agglomerations could be established. An analysis of whether there were norms with respect to high-tech and low-tech manufacturing agglomerations was carried out on groups of clusters of more or less creative regions.

b) Results

b.1) Results of the Correlations

The results of the correlations are shown in Table 3. To analyze them, we focused first on the cor-

Exhibit 1. LQ formula

$$LQd = \frac{\text{Firms in NACE}_i \text{ in region A / Total firms in NACE}_i \text{ in EU}}{\text{Manuf. or service firms in region A / Total manuf. or service firms in EU}}$$

Table 3. Descriptive statistics and Spearman's correlation coefficients (number of industry agglomerations and GDP)

	Mean	SD	High (LQs)	MedHigh (LQs)	MedLow (LQs)	Low non-creative (LQs)	HTKIS non-creative (LQs)	OKIS non-creative (LQs)	LKIS (LQs)	Creative (LQs)	GDP perinhab (€)
High (LQs)	0.92	0.84	1								
MedHigh (LQs)	2.31	1.60	0.649**	1							
MedLow (LQs)	2.69	1.23	0.249**	0.392**	1						
Low non-cre-ative[(1)] (LQs)	2.62	1.20	-0.262**	-0.244**	-0.050	1					
HTKIS non-creative[(2)] (LQs)	0.85	0.64	0.326**	0.208**	0.218**	-0.142*	1				
OKIS non-creative[(3)] (LQs)	2.85	1.88	0.613**	0.636**	0.282**	-0.223**	0.271**	1			
LKIS[(4)] (LQs)	5.27	2.10	0.283**	0.460**	0.323**	0.004	0.243**	0.578**	1		
Creative (LQs)	3.58	2.38	0.543**	0.442**	0.248**	-0.230**	0.455**	0.544**	0.219**	1	
GDP perinhab(€)	24,250	9,000	0.433**	0.455**	-0.011	-0.179**	0.105	0.419**	0.203**	0.440**	1

**Correlation is significant at the 0.01 level (bilateral). *Correlation is significant at the 0.05 level (bilateral).

High: High-tech manufacturing industries; MedHigh: Medium-high-tech manufacturing industries; MedLow: Medium-low-tech manufacturing industries; Low non-creative: Low-tech non-creative manufacturing industries; HTKIS non-creative: High-tech knowledge-intensive non creative services; OKIS non-creative: Other knowledge-intensive non-creative services; LKIS: Less-knowledge-intensive services; Creative: Creative industries; GDPperinhab: GDP per inhabitant

(1) Excluding wearing apparel, leather, and printing, included in "creative industries."

(2) Includes only telecommunications and information service activities as the rest (motion picture, video and television, sound recording and music, broadcasting, computer programming, and scientific research and development) are included in "creative industries."

(3) Excluding publishing, architectural and engineering activities, advertising, and arts, entertainment and recreation, included in "creative industries."

(4) Excluding retail sale of other goods in specialized stores, included in "creative industries."

Source: Eurostat. *Structural Business Statistics* (SBS) Database relations between sectors without taking GDP into account. Afterwards we looked at the correlations between GDP and industry (manufacturing and services).

From the results in Table 3, the following can be deduced:

- The results of the relationship between manufacturing sectors show that there is a positive correlation between high-tech manufacturing, medium-high manufacturing and medium-low manufacturing sectors. However, the correlation is negative between high-tech and low-tech non-creative sectors, and between medium-high and low-tech non-creative sectors. Consequently, the results do coincide with those of Heidenreich (2009), who concluded that high and medium-high coexist with medium-low manufacturing. Robertson and Patel (2007) also found this coexistence.
- High-tech, medium-high and medium-low manufacturing and the three groups of non-creative services are correlated. Therefore, contrary to what Heidenreich (2009) states, regions do not specialize in manufacturing or in services. Moreover, it is important to notice that low-tech non-creative manufacturing has a negative correlation with knowledge-intensive non-creative services.
- In the case of creative industries, there is a positive correlation with services as well as with manufacturing, except for low-tech non-creative manufacturing.

Based on these initial results, if there is a positive correlation with high-tech industries and knowledge intensive services, then it can be concluded that there are more high-tech industries in creative regions than other types of industries.

If we include the results obtained when relating sectors with GDP per inhabitant, the results obtained show:

- There is a positive correlation between GDP per inhabitant and the localization of high-tech manufacturing industries, medium-high tech manufacturing, OKIS non-creative (other knowledge-intensive non-creative services), and less-knowledge intensive services and creative industries.
- The correlation is negative with low-tech non creative industries.

These results do not coincide with Heidenreich (2009) or Vence-Deza and González-López (2008), who find that the wealth of a region does depend exclusively on high-tech industries and knowledge-intensive services.

Another result shows that there is a positive correlation between GDP per inhabitant and creative industries, which include low-tech manufacturing, HTKIS and OKIS. However, they do not include high-tech manufacturing industries. Therefore, creative regions do not depend exclusively on high-tech industries and knowledge-intensive services, contrary to the findings of Heidenreich (2009), Vence-Deza and González-López (2008).

b.2) Cluster Results

The cluster technique was conducted in order to obtain group structures. Thus, the process was divided into two stages, as recommended by various authors (Punj and Stewart, 1982; Hair et al., 1996:515). In the first phase, a hierarchic method was used to obtain a suitable number of groups, followed by a non-hierarchic method (k-means) to establish the group distribution and definitive composition, based on the number of groups obtained during the first stage. A dendogram was obtained for application in three to five groups. However, in our opinion, it seemed more suitable to choose five groups to show all the diversity

Table 4. Mean of agglomerations in the 5 clusters (Industry agglomerations and GDP PPS per inhabitant)

Agglomerations/ Clusters	1 High-manufacturing regions	2 LKIS regions	3 Intermediate regions	4 Low-tech non-creative regions	5 Super-creative regions
High-tech (LQs)	**2**	0	1	0	**2**
MedHigh-tech (LQs)	**4**	1	1	1	**4**
MedLow-tech (LQs)	3	2	2	2	3
Low-tech non creative (LQs)	2	3	2	4	2
KIHTS non creative (LQs)	1	1	1	1	1
OKIS non creative (LQs)	5	2	2	1	4
LKIS (LQs)	7	6	3	3	6
Creative (LQs)	**4**	1	4	2	**8**
Mean GDP PPS per inhabitant	26,180 (2nd)	22,900 (4th)	23,612 (3rd)	16,780 (5th)	**33,850 (1st)**
Number of regions	69	62	42	44	32

within the European regions. The coincidence between the two methods in the assignment of regions to the groups was 90%. Description of the groups is shown in Table 4.

In addition, the ANOVA analysis applied to each independent variable (LQs) used to obtain the groups revealed that all variables discriminate the classification into the five groups with all of them being significant at $p<0.01$. Additional statistical procedures were based on the ANOVA test to identify the differences between the group of belonging of regions and their GDP pps. The test was significant at $p<0.01$. This exercise shows that the group of belonging or cluster of regions, presents GDP pps differences across regions in general (Table 4). The results we obtained (Table 4) are:

- Five cases can be verified in the clusters depending on whether or not regions are creative, and the importance of all groups of industries. The first are the regions we call *high-manufacturing regions*; the second are the LKIS regions (*less-knowledge intensive services*); the third are the *intermediate regions*; the fourth are the *low-tech non-creative regions*; and the fifth are the *super-creative regions*, which have a higher mean of creative agglomerations than the rest.
- When calculating the mean GDP per inhabitant for these regions we can see that it is highest in the super-creative regions.
- High-manufacturing regions have a mean of agglomerations in all groups of industries which is similar than the super-creative, except for the creative industry group. Both regions have a higher number of OKIS non-creative agglomerations than the others.
- Creative agglomerations in intermediate regions are higher than those in LKIS regions and low-tech non-creative regions.

In super-creative regions there is a greater number of high-tech manufacturing industries and creative industries. However, the mean of low-tech industry agglomerations is very similar in the five clusters. Therefore, we can conclude that in the most creative regions, where the mean GDP per inhabitant is higher, high-tech manufacturing industries go hand in hand with low-tech manufacturing industries, which is comparable to the results obtained by Robertson and Patel (2007).

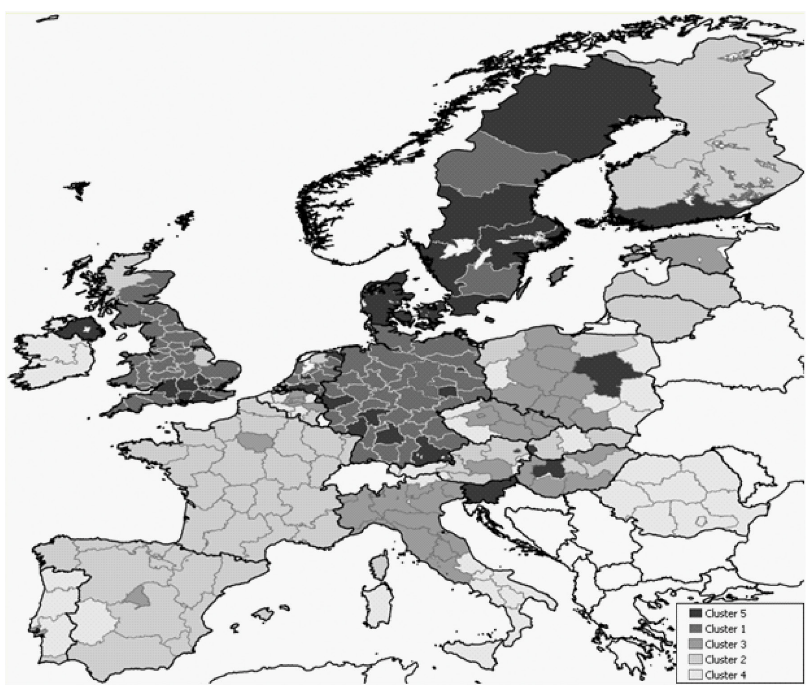
Examples of *super-creative regions* (Figure 1) include Berlin, Hamburg, Utrecht, Northern Holland, Vienna, Stockholm, Inner London, Outer London, Berkshire, Buckinghamshire and Oxfordshire. *High-manufacturing* regions include Düsseldorf, Greater Manchester, East Anglia and Essex. *Intermediate regions* include Madrid, Île de France, Veneto, Lazio and Emilia-Romagna. *LKIS regions* include Catalonia, Valencian Region, Salzburg and Brittany. *Low-tech non-creative regions* include Puglia, Campania and Extremadura.

5. CONCLUSION

This paper strives to answer the question of how much influence the existence of manufacturing, service and creative industrial agglomerations has on a region's wealth. To respond to this question, the industrial agglomerations of manufacturing, services and creative industries of 250 regions in 24 European countries were calculated. Afterwards, the relationships between agglomerations and between agglomerations and GDP were verified.

The studies carried out up until now on the maps of manufacturing and service agglomerations (Leydesdorff and Fritsch 2006, Leydesdorff et al 2006, Vence-Deza and González-López 2008, Heidenreich 2009) have shown the importance of the relationships between both. However, when analyzing whether the most important sectors in the development of a region are high-tech and services, as measured by GDP, the results obtained do not always coincide (Vence-Deza and González-López 2008, Leydesdorff and Fritsch 2006, Leydesdorff et al 2006). Additionally, the studies usually show an incompatibility between the agglomerations of high-tech and low-tech manufacturing sectors within the same region (Heidenreich 2009). Nevertheless, the empirical analysis carried out in this paper demonstrates that this incompatibility should be qualified. In

Figure 1. Regions in the five clusters



fact, in the most creative regions the agglomerations of high-tech manufacturing industries are greater than in the rest of the regions, while they are similar in low-tech manufacturing industries. The compatibility of high-tech and low-tech manufacturing industries was also shown by Robertson and Patel (2007).

In addition, this paper verifies that the wealth of a region is not only related to the agglomerations of high-tech manufacturing industries found in the region, but also to the creative industries found there. Another relationship which has been verified in this paper is the relationship of creative industries and knowledge-intensive services (KIS), which emphasizes their importance in the creation of knowledge and regional development (Windrum and Tomlinson, 1999; Bishop, 2008; Aslesen and Isaksen, 2007a; Strambach, 2008).

Three important conclusions can be inferred from the results obtained in this paper. The first is that creative industries play an important role in the wealth of a region. The second is that the most creative regions are characterized by having more high-tech manufacturing industries than the rest of the regions although the number of low-tech manufacturing industries is similar. Lastly, in the richest regions a greater number of high-tech manufacturing industries coexist with a greater number of creative industries.

The contributions and results found in this work are important for both academia as well as for policymakers. It opens new lines of research for the former in the relationships between industries, as our study goes beyond those carried out on manufacturing and services. Policymakers will find the study of use because the results show the role creative industries play on regional wealth, in addition to demonstrating that the most creative regions have a need for the coexistence of diversified sectors that simultaneously include a high share of both high-tech manufacturing and creative industries.

ACKNOWLEDGEMENT

The authors would like to thank Ministry of Science and Innovation and Universitat Politècnica de València (Spain) for financially supporting this research (ECO2010-17318 MICINN Project and Research Project n. 2677-UPV).

REFERENCES

Aslesen, H. W., & Isaksen, A. (2007a). New perspectives on knowledge-intensive services and innovation. *Geografiska Annaler. Series B, Human Geography*, *89*, 45–58. doi:10.1111/j.1468-0467.2007.00259.x

Aslesen, H. W., & Isaksen, A. (2007b). Knowledge intensive business services and urban industrial development. *The Service Industries Journal*, *27*(3), 321–338. doi:10.1080/02642060701207239

Baum, S., O'Connor, K., & Yigitcanlar, T. (2009). The implications of creative industries for regional outcomes. *International Journal of Foresight and Innovation Policy*, *5*(1-3), 44–64. doi:10.1504/IJFIP.2009.022098

Becattini, G., & Coltorti, F. (2006). Areas of large enterprise and industrial districts in the development of post-war Italy: A preliminary survey. *European Planning Studies*, *14*(8), 1105–1138. doi:10.1080/09654310600852415

Bishop, P. (2008). Spatial spillovers and the growth of knowledge intensive services. *Tijdschrift voor Economische en Sociale Geografie*, *99*(3), 281–292. doi:10.1111/j.1467-9663.2008.00461.x

Boix, R., & Trullén, J. (2011). La relevancia empírica de los distritos industriales marshallianos y los sistemas productivos locales manufactureros de gran empresa en España. *Investigaciones Regionales*, *19*, 75–96.

Capone, F. (2008). Mapping and analysing creative systems in Italy (1991-2001) In Cooke, P., & Lazzeretti, L. (Eds.), *Creative cities, cultural clusters and local economic development* (pp. 338–364). Cheltenham, UK: Edward Elgar.

Center for Strategy and Competitiveness. (2009). *Priority sector report: Knowledge intensive business services*. Retrieved from www.clusterobservatory.eu

Cooke, P. (2008). Culture, clusters, districts and quarters: Some reflections on the scale question In Cooke, P., & Lazzeretti, L. (Eds.), *Creative cities, cultural clusters and local economic development* (pp. 25–47). Edward Elgar.

Cooke, P., De Laurentis, C., Tödtling, F., & Trippl, M. (2007). *Regional knowledge economies. Markets, clusters and innovation*. Cheltenham, UK: Edward Elgar.

Crouch, C., & Farrell, H. (2001). Great Britain: Falling through the holes in the network concept In Crouch, C. (Eds.), *Local production systems in Europe: Rise or demise?* (pp. 154–211). Oxford, UK: Oxford University Press.

DCMS. (2009). *Creative industries economic estimates statistical bulletin.*

De Propris, L., Chapain, C., Cooke, P., MacNeill, S., & Mateos-García, J. (2009). The geography of creativity. *NESTA Interim Report*, August.

Doloreux, D., Amara, N., & Landry, R. (2008). Mapping regional and sectoral characteristics of knowledge-intensive business services: Evidence from the Province of Quebec (Canada). *Growth and Change*, *39*(3), 464–496. doi:10.1111/j.1468-2257.2008.00434.x

Drejer, I., & Vinding, A. L. (2005). Location and collaboration: Manufacturing firms' use of knowledge intensive services in product innovation. *European Planning Studies*, *13*(6), 879–898. doi:10.1080/09654310500188407

European Affairs, K. E. A. (2006). *The economy of culture in Europe*. Brussels, Belgium: European Commission Directorate - General for Education and Culture.

Garnham, N. (2005). From cultural to creative industries: An analysis of the implications of the creative industries approach to arts and media policy making in the United Kingdom. *International Journal of Cultural Policy*, *11*(1), 15–29. doi:10.1080/10286630500067606

Hair, J., Anderson, R., Tatham, R., & Black, W. (1996). *Análsis Multivariante* (5th ed.). Madrid, Spain: Prentice Hall.

Hatzichronoglou, T. (1997). *Revision of the high-technology sector and product classification*. OECD Science, Technology and Industry Working Papers, 1997/2, OECD Publishing.

Healy, K. (2002). What's new for culture in the new economy? *Journal of Arts Management Law and Society*, *32*(2), 86–103.

Heidenreich, M. (2009). Innovation patterns and location of European low- and medium-technology industries. *Research Policy*, *38*(3), 483–494. doi:10.1016/j.respol.2008.10.005

Howkins, J. (2005). The mayor's commission on the creative industries In Harteley, J. (Ed.), *Creative industries* (pp. 117–125). London, UK: Blackwell.

Larsen, J. N. (2001). Knowledge, human resources and social practice: The knowledge-intensive business service firm as a distributed knowledge system. *The Service Industries Journal*, *21*(1), 81–102. doi:10.1080/714004998

Lazzeretti, L., Boix, R., & Capone, F. (2008). Do creative industries cluster? Mapping creative local production systems in Italy and Spain. *Industry and Innovation*, *15*(5), 549–567. doi:10.1080/13662710802374161

Lazzeretti, L., Boix, R., & Capone, F. (2009). *Why do creative industries cluster? An analysis of the determinants of clustering of creative industries*. Paper presented at DRUID Summer Conference 2009.

Leydesdorff, L., Dolfsma, W., & Van der Panne, G. (2006). Measuring the knowledge base of an economy in terms of triple-helix relations among "technology, organization and territory." *Research Policy*, *35*(2), 181–199. doi:10.1016/j.respol.2005.09.001

Leydesdorff, L., & Fritsch, M. (2006). Measuring the knowledge base of regional innovation system in Germany in terms of a Triple Helix dynamics. *Research Policy*, *35*(10), 1538–1553. doi:10.1016/j.respol.2006.09.027

Lorenzen, M., & Frederiksen, L. (2008). Why do cultural industries cluster? Localization, urbanization, products and projects In Cooke, P., & Lazzeretti, L. (Eds.), *Creative cities, cultural clusters and local economic development* (pp. 155–179). Cheltenham, UK: Edward Elgar.

Miles, I. (2008). Knowledge services In Hearn, G., & Rooney, D. (Eds.), *Knowledge policy. Challenges for the 21st century* (pp. 11–26). Cheltenham, UK: Edward Elgar.

Muller, E., & Doloreux, D. (2009). What we should know about knowledge-intensive business services. *Technology in Society*, *31*(1), 64–72. doi:10.1016/j.techsoc.2008.10.001

Nählinder, J. (2005). *Innovation and employment in services. The case of knowledge business services in Sweden*. Linköping, Sweden: Unitryck Linköping.

O'Connor, J. (2007). *The cultural and creative industries: A review of the literature*. Creative Partnerships Series. Arts Council England.

O'Donoghue, D., & Gleave, B. (2004). A note on methods for measuring industrial agglomeration. *Regional Studies*, *38*(4), 419–427. doi:10.1080/03434002000213932

Peneder, M., Kaniovski, S., & Dachs, B. (2003). What follows tertiarisation? Structural change and the role of knowledge-based services. *The Service Industries Journal*, *23*(2), 47–66. doi:10.1080/02642060412331300882

Pilat, D., & Wölfl, A. (2005). *Measuring the interaction between manufacturing and services*. OECD Science, Technology and Industry Working Papers, 2005/5, OECD Publishing.

Pitelis, C., & Pseiridis, A. (2006). A conceptual framework for firm cooperation and clusters, and their impact on productivity In Pitelis, C. (Eds.), *Clusters and globalisation. The development of urban and regional economies* (pp. 17–60). Cheltenham, UK: Edward Elgar.

Power, D. (2002). Cultural industries in Sweden: An assessment of their place in the Swedish economy. *Economic Geography*, *78*, 103–127. doi:10.2307/4140784

Power, D. (2003). The Nordic cultural industries: a cross-national assessment of the place of the cultural industries in Denmark, Finland, Norway and Sweden. *Geografiska Annaler. Series B, Human Geography*, *85*(3), 167–180. doi:10.1111/j.0435-3684.2003.00139.x

Power, D., & Nielsén, T. (2010). *Priority sector report: Creative and cultural industries*. European Commission, Europe Innova -European Cluster Observatory.

Pratt, A. C. (2005). Cultural industries and public policy: An oxymoron? *International Journal of Cultural Policy*, *11*(1), 31–44. doi:10.1080/10286630500067739

Pratt, A. C. (2008). Creative cities: The cultural industries and the creative class. *Geografiska Annaler. Series B, Human Geography*, *90*(2), 107–117. doi:10.1111/j.1468-0467.2008.00281.x

Punj, G., & Stewart, D. (1983). Cluster analysis in marketing research: Review and suggestions for application. *JMR, Journal of Marketing Research*, *20*, 134–148. doi:10.2307/3151680

Robertson, P. L., & Patel, P. R. (2007). New wines in old bottles: Technological diffusion in developed economies. *Research Policy*, *36*(5), 708–721. doi:10.1016/j.respol.2007.01.008

Shearmur, R., & Doloreux, D. (2009). Place, space and distance: Towards a geography of knowledge-intensive business services innovation. *Industry and Innovation*, *16*(1), 79–102. doi:10.1080/13662710902728001

Stam, E., de Jong, J. P. J., & Marlet, G. (2008). Creative industries in The Netherlands: Structure, development, innovativeness and effects on urban growth. *Geografiska Annaler. Series B, Human Geography*, *90*(2), 119–132. doi:10.1111/j.1468-0467.2008.00282.x

Strambach, S. (2008). Knowledge-intensive business services (KIBS) as drivers of multilevel knowledge dynamics. *International Journal of Services Technology and Management*, *10*(2-4), 152–174. doi:10.1504/IJSTM.2008.022117

Towse, R. (2010). Creativity, copyright and the creative industries paradigm. *Kyklos*, *63*(3), 461–478. doi:10.1111/j.1467-6435.2010.00483.x

UNCTAD. (2010). *Creative economy. Report 2010*. Geneva: UNDP-UNCTAD.

Vence-Deza, X., & González-López, M. (2008). Regional concentration of the knowledge-based economy in the EU: Towards a renewed oligocentric model? *European Planning Studies*, *16*(4), 557–578. doi:10.1080/09654310801983472

Windrum, P., & Tomlinson, M. (1999). Knowledge-intensive services and international competitiveness: A tour country comparison. *Technology Analysis and Strategic Management*, *11*(3), 391–408. doi:10.1080/095373299107429

Wood, P. (2006). The regional significance of knowledge-intensive services in Europe. *Innovation (Abingdon)*, *19*(1), 51–66. doi:10.1080/13511610600607718

This work was previously published in Comparing High Technology Firms in Developed and Developing Countries: Cluster Growth Initiatives, edited by Tomas Gabriel Bas and Jingyuan Zhao, pp. 65-79, copyright 2012 by Inormation Science Reference (an imprint of IGI Global).

Chapter 92
Design of Manufacturing Cells Based on Graph Theory

José Francisco Ferreira Ribeiro
University of São Paulo, Brazil

ABSTRACT

In this chapter a comparative study is presented between (I) sequential heuristics, (II) simulated annealing, (III) tabu search, and (IV) threshold algorithm for graph coloring and its application for solving the problem of the design of manufacturing cells in a job shop system production. The job shop production system has a very large proportion of all manufacturing activity. The principal concepts of manufacturing cells, graph theory, and heuristics are presented. The results obtained with these algorithms on several examples found in the literature are consistently equivalent with the best solution hitherto known in terms of numbers of inter-cell moves and dimensions of cells.

INTRODUCTION

Over 75% of all parts manufactured in the industry are produced in batches of 50 parts or less. Consequently, the production in batch and the production on demand constitute a considerable proportion of all manufacturing activity (Groover, 1987). Job shop is a production environment that produces parts in small batches. It's a production environment common in small and medium enterprises. The parts require different manufacturing operations and must be performed through various production departments and in different sequences (Oliveira, Ribeiro and Seok, 2009). Orders differ in the number of parts, design, processing times, setup times or urgency. The high demand for machinery and the different production sequences can cause long queues in the shop floor. The consequence is delivery times unreliable, whereas nowadays delivery times should be short and reliable (Ribeiro and Pradin, 1993). Clustering is the task of classifying a collection of objects, such as documents, parts, or machines, into natural categories. Clustering techniques are widely

DOI: 10.4018/978-1-4666-1945-6.ch092

used in many areas such as machine learning, data mining, pattern recognition, image analysis and bioinformatics. The design of manufacturing cells or clustering consists in partitioning the set of parts to be manufactured in an industry into families and the available machines into groups or cells, so that each family is associated with one machine group, and vice-versa. Each family-group pair constitutes a manufacturing cell. This concept lies on grouping similar parts in families, proposing to produce them in cells that have specially selected machines to accomplish this. This procedure leads to greater automation, set up time reduction, standardization of the tools used and a reduction of manufacturing cycles (Ribeiro and Meguelati, 2002). A greater efficiency in management and manufacturing is expected, due to the decomposition of the production global system in sub-systems of reduced dimension. Workshops operating on this principle, called Group Technology (Burbidge, 1975), offer a reduction in unproductive manufacturing time, resulting in greater flexibility, just in time and productivity (Hyer and Wemmerlöw, 1989).

However, the design of manufacturing cells requires the solution of a complex mathematical problem: the block-diagonalization of an incidence matrix [parts × machines] corresponding to the global production system. This block-diagonalization uses as its optimization criterion the minimization of the number of elements in the matrix outside the diagonal blocks. These elements represent inter-cell moves and, in practice, imply undesirable movement of parts to machines in other cells that are not present in the cell to which the part is assigned. That is the reason why, regarding the manufacturing cells, there is an attempt to minimize the number of inter cell moves, at the same time that a balance of workloads between the different cells projected is sought. This is treated as a combinatorial problem for which there is no polynomial time algorithm (Garey and Johnson, 1979), and the most common approach in the literature is to propose heuristic algorithms.

BACKGROUND

A large number of techniques have been used in recent years (Sing, 1993) to hit the block-diagonalization of the matrix, designing manufacturing cells and implementing Group Technology in the industries. For example:

a. Mathematical programming: Won (2000), Albadawi, Bashir, and Chen (2005), Panchalavarapu and Chankong (2005), Slomp, Chowdary, and Suresh (2005), Rajagopalan and Fonseca (2005), Adil and Ghosh (2005), Yin, Yasuda, and Hu (2005), Foulds, French, and Wilson (2006)
b. Branch and bound: Ramabhatta e Nagi (1998), Boulif and Atif (2006)
c. Fuzzy logic: Xu e Wang (1989), Chu and Hayya (1991)
d. Genetic algorithms: Suer, Vásquez, and Peña (1999), Zhao and Wu (2000), Dimopoulos and Mort (2001), Suer, Peña, and Vázquez (2003), Gonçalves and Resende (2004), Hicks (2004), Solimanpur, Vrat, and Shankar (2004), Rajagopalan and Fonseca (2005), Vin, De Lit, and Delchambre (2005), Gonçalves and Tiberti (2006), Jeon and Leep (2006)
e. Neural networks: Lozano, Canca, Guerrero, and Garcia (2001), Guerrero, Lozano, Smith, Canca, and Kwok (2002), Solimanpur, Vrat, and Shankar (2004), Pashkevich and Kazheunikau (2005)
f. Meta-heuristics — tabu search and simulated annealing: Caux, Bruniaux, and Pierreval (2000), Baykasoglu (2003), Spiliopoulos and Sofianopoulou (2003), Xambre and Vilarinho (2003), Cao and Chen (2005)
g. Data analysis: Ribeiro and Pradin (1993), Diallo, Pierreval, and Quilliot (2001), Rios, Campbell, and Irani (2002), Ribeiro and Meguelati (2002), Ribeiro (2003), Oliveira, Ribeiro and Seok (2009)

h. Graph theory: Rajagopalan and Batra (1975), Askin and Chiu (1990), Rath, Das and Sahu (1995), Selim (2002), Ribeiro (2009)

The main different methods from the literature are approximate methods (not optimal), because the problem is NP-complete. The computational time is very high when we use exact methods to solve large-size examples.

GRAPH COLORING

A coloring of a graph G(V, A), with), with V a set of N nodes and A a set of M edges (Matula, Marble, and Isaacson, 1972, Christofides, 1975, Korman, 1979) is an assignment of any color belonging to set of colors C = {ci} for each node of V, where two nodes connected by an edge of A cannot have the same color.

I.e., a coloring of G is a function f: V → C | if (i, j) ∈ A ⇒ f(i) ≠ f(j). A k-coloring of G(V, A) is a coloring with k colors, i.e., a partition of V in k independent sets of nodes.

In this case, G is a k-coloring graph. The chromatic number λ(G) of G is the smaller number of colors for which a k-coloring of G exists. The graph coloring problem is NP-complete (Garey and Johnson, 1979).

It is a high combinatorial problem: for example, a complete and exhaustive enumeration of all colorings consists in an $O(M.k^N)$ algorithm (Aho, Hopcroft and Ullman, 1983). But, some immediate results can be obtained of the definition, such as:

a. A graph G is bi-chromatic if and only if is bipartite
b. A complete sub-graph of G with t vertices Kt is t-chromatic
c. A graph G with two or more edges is at least 2-chromatic

The concepts of graph coloring, clique (a complete sub-graph of G) and independent set of nodes are close:

a. k colors are necessary to coloring k nodes of a clique with cardinality k, thus λ(G) is greater than or equal to the cardinality of the greatest clique of G
b. Let us consider $S_1 \dots S_k$ disjunctive subsets of S generated by the k-coloring; then = $\cup S_i$ = S, i = 1 … k, and each S_i is an independent set and each Si is an independent set in which any couple of edges is not connected.

GRAPH COLORING AND MANUFACTURING CELLS DESIGN

In a graph G(V, A) associated to a production system, V = {set of parts to be manufactured}. The dissimilarity indexes between parts are calculated, for example, based on the existing differences between their production sequences.

In this case, an edge connecting two nodes i and j exists if the dissimilarity dij is greater than or equal to a critical dissimilarity established a priori.

This critical dissimilarity can be modified in any instant with the objective of higher or lower number of edges of the graph and then, the number designed of cells.

Two nodes connected cannot have the same color, resulting in a coloring where the parts with different colors have dissimilarity indexes greater than or equal the fixed value for critical dissimilarity.

Parts with the same color will be assigned to the same family and manufactured in the same cell of machines.

The distribution of machines is carried out considering the number of operations carried out in each family: a machine is assigned to the family in which it will be most used.

Table 1. Matrix MPM-LOAD [parts × machines]

	Machine 1	Machine 2	Machine 3	Machine 4
Part 1	0	15	0	40
Part 2	25	0	60	0
Part 3	0	35	0	50
Part 4	90	0	70	0
Part 5	30	0	0	0

Example

From the initial data a matrix MPM-LOAD[parts × machines] is obtained (see Table 1), which provides the total time spent by each part transiting in the program of each machine. This matrix is called work load matrix and its coefficients are calculated as follows:

load[i,j] = unit[i] × ∑ duration[i, k]
k | program[i, j] = j

where:

unit[i] = unit number of parts[i] to manufacture.
duration[i, k] = duration of operation k on a part[i].
program[i, k] = type of machine used to perform operation k on a part[i].

Let the matrix MPM [parts × machines] (see Table 2), that informs the use ($MPM_{ij} = 1$) or not ($MPM_{ij} = 0$) of part i by the machine j. The matrix D [parts × parts] (see Table 3) gives the dissimilarity index between parts ("distances between parts") to be manufactured. Two manufacturing cells will be designed for this workshop. If the critical dissimilarity is equal to 1, G(V, A) is the graph shown in Figure 1, i.e., three-coloring, and not acceptable because the workshop must be partitioned in two cells. Then, the critical dissimilarity value is incremented to two and the two-coloring graph shown in Figure 2 is obtained. After defining the families of parts, machines are assigned to the most demanding families. The distribution of machines is carried out considering the number of operations carried out in each family: a machine is assigned to the family in which it will be most used. The matrix CELLS [parts × machines] in Table 4 presents the workshop partitioning in 2 manufacturing cells and this solution is very good: there are no inter-cell moves and the dimensions of cells are equal to [2 × 2] for cell 1 and [3 × 2] for cell 2.

DISSIMILARITY INDEXES BETWEEN PARTS

Similarity measures for parts or machines have a long history of use for manufacturing cells design

Table 2. Matrix MPM [parts × machines]

	Machine 1	Machine 2	Machine 3	Machine 4
Part 1	0	1	0	1
Part 2	1	0	1	0
Part 3	0	1	0	1
Part 4	1	0	1	0
Part 5	1	0	0	0

Table 3. Matrix D [parts × parts]

	Part 1	Part 2	Part 3	Part 4	Part 5
Part 1	—				
Part 2	4	—			
Part 3	0	4	—		
Part 4	4	0	4	—	
Part 5	3	1	3	1	—

Table 4. Matrix CELLS [parts × machines]

	Machine 2	Machine 4	Machine 1	Machine 3
Part 1	1	1		
Part 3	1	1		
Part 2			1	1
Part 4			1	1
Part 5			1	

Figure 1. Three-coloring graph

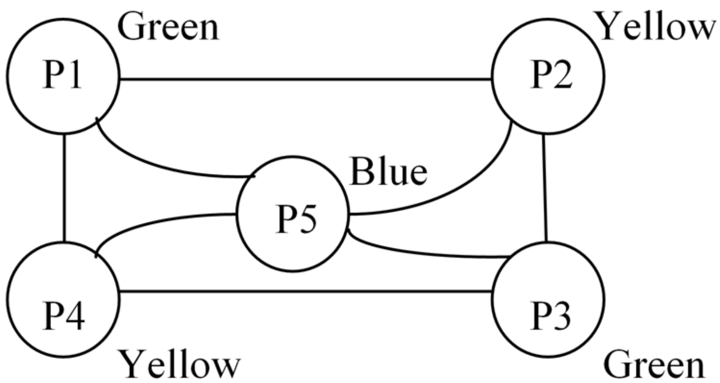
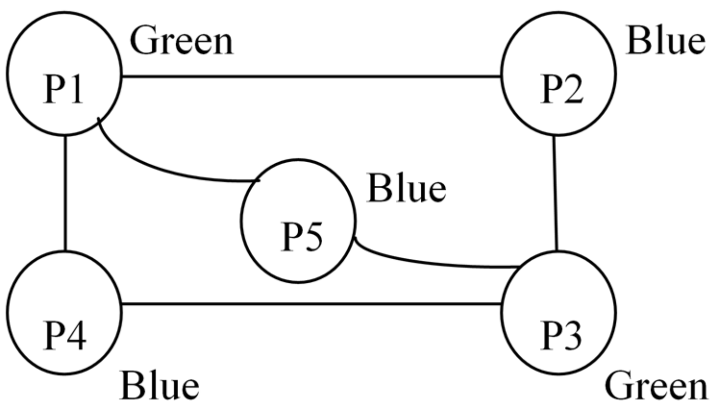


Figure 2. Two-coloring graph



problems. The first use of similarity measure for the manufacturing cells design problem was by McAuley (1972). Let:

x_{ij} = number of parts processed by machines mi and m_j (number of matches)
x_i = number of parts processed by machine m_i only
y_{ij} = number of parts that are not processed by either m_i or m_j (number of misses).

McAuley (1972) adopted a Jaccardian similarity measure in evaluating similarities of pairs of machines. This measure considers x_{ij} as a main factor for similarity and divides it by the number of parts which either machine m_i or m_j processes. After the use of the first similarity measure, many other different types of similarity measures have been introduced, for example: Bipartite, Kusiak, McAuley-Jaccard, BUB, Russel and Rao, Sorenson, Simple Matching, Ochiai (Oliveira, Ribeiro and Seok, 2008). Jaccardian similarity measures (Bipartite, McAuley-Jaccard, Sorenson and Ochiai) are adopted for many cell formation problems (Wei and Kern, 1989). Sorenson's measure is a simple modification of McAuley's measure, where xij has more weight (Romesburg, 1984). However, it is reported that these Jaccardian similarity measures are limited because they do not use the number of misses, yij (Moiser, 1989, Islam and Sarker, 2000). Selim and Abdel (2003) distinguished those Jaccardian similarity measures from non-Jaccardian measures which use yij (Kusiak, BUB, Russel and

Rao, Simple Matching). Among them, Kusiak's measure is one of the simplest (Kusiak, 1987). As said before, they use not only xij but also yij in the similarities. Simple matching was first introduced in the field of medicine. This measure simply adds yij to both the top and bottom of McAuley's Jaccardian measure. The most successful non-Jarccardian measure is the BUB similarity measure. This measure has been successfully adopted by many clustering algorithms because their distribution of values are more normal and continuous (Baroni-Urban and Buser, 1976). Other notable non-Jarccardian measures are Russel and Rao's and Ochiai's in Romesburg (1984). Note that there are two big differences between the bipartite similarity matrices and other similarity coefficients usually used for manufacturing cells design problems. First, while we construct both similarity matrices for machines and parts, most array-based clustering algorithms use only a machine-machine similarity matrix to construct cells of closely related machines and then part families are constructed according to these machine cells. Second, diagonal entries of similarity coefficient matrices do not play any role, so all diagonal entries are considered 0. However, it is reasonable when the similarity of each machine or part with itself has bigger similarity than with others.

More details on the similarity measures can be found in Sarker (1996) and Sarker and Islam (1999) that are dedicated to analyzing most of the existing similarity measures in the literature based on a set of important properties developed by Baroni-Urban and Buser (1976).

In this chapter the dissimilarity indexes between parts i and j are calculated by quantifying the differences between their production sequences. The proposed method compares the production sequences of parts i and j utilizing the binary matrix MAT [parts × machines], where:

$$\mathrm{MAT}_{ij} = \begin{cases} 1 & \text{if machine j is utilized to manufacture part i} \\ 0 & \text{if machine j is not utilized to manufacture part i} \end{cases}$$

Then, the dissimilarity index between parts i and k is computed as follows:

$$D[i,k] = \sum_{j=1}^{\text{number of machines}} \boldsymbol{\mu}[\mathrm{MAT}_{ij}, \mathrm{MAT}_{kj}]$$

where:

$$\mu[a_{ij}; a_{kj}] = \begin{cases} 1 & \text{if } \mathrm{MAT}_{ij} \neq \mathrm{MAT}_{ij} \\ 0 & \text{otherwise} \end{cases}$$

This procedure is based on the similarity computation used, for example, by Kusiak (1987), Wei and Kern (1989), Ribeiro and Meguelati (2002), Oliveira, Ribeiro and Seok (2008) and Ribeiro (2009).

ALGORITHMS

Sequential Heuristics

Let us consider a Table that defines an order for the N nodes of G. A sequential method S consists in coloring successively $V_1 \dots V_N$, assigning to these nodes the color of the minimum index not utilized by their neighbors. The choice of the node is imposed in each iteration and the choice of the color is trivial, corresponding to an algorithm sample and compact in $O(N^2)$.

But, this heuristic can be very detrimental, despite the existence of an optimal order. In fact, if $\lambda(G) = k$, then a partition of nodes exists in k independent set of nodes $S_1 \dots S_k$. Inserting in V the nodes in crescent order of the number of the independent set of nodes, this heuristic finds the optimal solution searching N! orders. An order by the nodes can be established, for example, by three manners as follows:

a. The simpler manner to implement S consists in utilizing the natural order of the numbers of nodes, i.e., $V_i = i$ for all node i. This heuristic is called FFS (First Fit Sequential).

b. By utilizing the crescent order of the degree of the nodes, Wesh and Powell (1967) propose the algorithm called LFS (Largest First Sequential).

c. An order is defined as follows: I) V_N is the node with minimum degree; II) For i = N-1 … 1, the node V_i is the node of minimum degree in the sub-graph generated by $V - \{V_N \ldots V_{i+1}\}$. This order is the basis of the algorithm proposed by Matula, Marble, and Isaacson (1972) and it is known by SLS (Smallest Last Sequential).

These three heuristics obtain good and bad results at random. Then, the implementation of meta-heuristics simulated annealing and tabu search utilizes them for obtaining an initial solution.

Simulated Annealing

Kirkpatrick, Gelatt and Vecchi (1983) have created the simulated annealing meta-heuristic. This technique is based on the annealing of metals in metallurgy: a metal cooled very fast presents a lot of microscopic defeats, corresponding to a local minimum in optimization problems. If the cooling is slow, the atoms arrange its structure, the defeats missing and the metal has, then, an ordered structure, equivalent to the global minimum in optimization. Simulated annealing in combinatorial optimization has some similarity with thermodynamics. The energy of the system is represented by an arbitrary real number T, i.e., the temperature.

Simulated annealing begins by local search or heuristic procedure (FFS, LFS or SLS), because the method always begins from a feasible initial solution s, and the six steps bellow are activated:

a. Take at random a transformation s' instead to find the best or the first improved solution in the neighborhood

b. Build the resulting solution s' and calculate the variation of cost $\Delta f = f(s') - f(s)$

c. If $\Delta f \leq 0$ the cost is better. Make s = s'

d. If $\Delta f > 0$, the cost is worse. The penalty is higher when the temperature is low and Δf is big. An exponential function has the appropriate properties. A probability of acceptation $a = e^{-\Delta f/T}$ is calculated and, then a parameter p is taken at random in [0, 1]. If $p \leq a$, the transformation is accepted, even with degradation of the cost and the solution is modified: s = s'. Otherwise, the transformation is rejected and s is conserved for the next iteration

e. The convergence is assured by lowering T slowly at each iteration, for example, T= kT, k < 1, but close

f. The algorithm stops when T attempts a limit equal to a ε fixed close to 0. In the implementation, the neighborhood of the solutions is defined by lists, but not on the graph.

Tabu Search

Tabu search was created by Glover (1989, 1990) and has no stochastic character. For the same computational time, tabu search generally presents better results than simulated annealing. Three fundamental points constitute the technique:

a. In each iteration, the neighborhood V(s) of the actual solution s is completely examined and the best solution s' is chosen, even if the cost is higher

b. Tabu list T forgives the return to a solution recently obtained. This list stocks in a compact manner, the steps traversed by the algorithm. Then, the method seeks s' in V(s) – T

c. The best solution found is stocked, because unlike the simulated annealing, it is rarely the last one

The method stops after a maximum number of iterations or after a maximum number of iterations without improving the best solution or when V(s) – T = ∅. The problematic point is the capacity C of the tabu list T. Glover's research knows that C = 7 to 20 is sufficient to prevent cycles, for any dimension of the problem. T is utilized as a short term memory. In each iteration, the C^{th} transformation of T (the oldest) is substituted by the last transformation realized. In the implementation, T is simply generated as a data structure of type queue. Werra (1990) and Hertz (1991) propose tabu search to solve the sub-problem of the existence of the k-coloring, for k fixed a priori. The technique begins to assign k arbitrary colors to the N nodes. If the k-coloring is attempted, the coloring procedure is repeated for k-1 colors. Frequently, the assignment presents conflict, i.e., nodes connected with the same color. Then, the objective consists in minimizing the total number of conflicts, converting the problem of existence in an optimization problem on the number of conflicts. The neighborhood is formed by all assignments obtained by exchanging the color i of a node α to another color j (α must have at least one neighbor with color i). Tabu search forgives α to take its previous color during NT iterations, NT = capacity of tabu list. In the iteration, the search seeks the best neighbor of the actual assignment, i.e., the neighbor that minimizes the number of conflicts. The search stops if the number of conflicts is equal to 0 (it means a k-coloring) or if the maximal number of searches in the neighborhood is over. The method is considered good for finding a coloring close to the optimal. It begins with a value of k corresponding to a superior limit, obtained, for example, by a heuristic procedure. If the tabu search solves the problem with k colors, the procedure is repeated for k-1.

The results described in the literature are very impressive, even for big graphs (10,000 nodes = 10,000 parts). The initial number of iterations of the search is lower (0 to 10 iterations), but increase quite a bit for the last values of k. This fact shows the hardiness of the existence problem when the procedure is close to the optimal solution.

Threshold Algorithm

Provided that the problem of partitioning an X collection of n objects with a dissimilarity of classes, in a fixed k number, so that the diameter d(P) among the classes is minimum. This corresponds to constructing a threshold G_s graph, a partial sub-graph of G(V<A), that should be k-colorable. To color the graphs of the vertices, a technique based on the algorithm by Guénoche (1993) was used. This algorithm enumerates all of the partitions in a fixed number of minimum diameter classes. It is based on a threshold graph in p colors, with each color defining a class. Many heuristics to approximate the diameter and the partitions of minimum diameter are enumerated only at the last stage. Let d(P) the diameter of the partition of V into p classes. A superior limit, s, for d(P) is heuristically determined. Then, the k-colorations for G_s are enumerated. As long as there is at least one possible k-coloration, the value s is decreased and a new iteration is processed. When the algorithm stops, the highest value obtained is equal to the remaining partition diameters. To enumerate all of the possible partitions of X, in minimum diameter K classes, this algorithm is applied, which is comprised of 3 stages and has an $O(k^{N-k})$ complexity. In the first stage a heuristic method is used to determine the maximum s, highest approximation to d(P), such that Gs will be k-colorable. The maximum s is determined by a method of dichotomic subdivisions of the variation interval of the dissimilarities in a sequential coloring method, in this case the saturation algorithm (called "Dsatur"). In this algorithm, at each iteration, the rate of saturation $DS_i(\alpha)$ is defined

as the number of colors already employed by the neighbors of α. The procedure consists of:

a. coloring the most sizable vertex with the color 1
b. in the following stages, take the vertex free of maximum DS and color it with the least possible index color

In the second stage, once the maximum s has been set, all of the Gs colorings are enumerated in k colors. With the aid of "Dsatur", maximum if possible, a clique of G_s is obtained, where each node represents a class. Then, according the order of saturation, the other vertices are colored in all possible manners. If a vertex is near the colored vertices, these will not be able to be used to color them. Thus, all of the partitions in k classes of less than s diameter are obtained. In the third stage, a decreasing order of the edges of the dissimilarity values, starting from s, are considered. An edge can be inserted in the graph as long as there is a compatible partition, that is, if the edge connects different types of nodes.

Thus, each inserted edge eliminates some previously obtained partitions. The first edge that cannot be inserted, since there would be no more compatible partitions left, has equal value to the highest interclass dissimilarity value. The remaining partitions for a fixed number of classes are of minimum diameter.

RESULTS

Table 5 summarizes some tests carried out with the programs corresponding to the sequential heuristics, simulated annealing, tabu search and threshold techniques. These programs are written in MatLab and run on a microcomputer. Using examples found in the literature, a comparison is presented between the best solution known (BS) and the solutions obtained by sequential heuristics (SH), simulated annealing (SA), tabu search (TS) and threshold method (TM).

The result presented to the sequential heuristics is the best solution obtained by the three options described above. Parameters, such as temperature, ε, NT, etc., for running the programs corresponding to simulated annealing and tabu search algorithms were chosen differently for each example, by attempt, always with the objective to find the best solution. In the 1st column of Table 5, the example given is given; in the 2nd, the methods utilized for solving the example: BS (the method responsible by the best solution known), SH: sequential heuristics, SA: simulated annealing, TS: tabu search, TM: threshold method; in the 3rd, the number of cells; in the 4th, the number of parts and machines; in the 5th, the number of inter-cell moves; in the 6th, the dimensions of the cells obtained; in the 7th, the computational time (in seconds — Pentium 2.20 GHz, 1.99 GB RAM).

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this chapter, a comparison between sequential heuristics (SH), meta-heuristics simulated annealing (SA) and tabu search (TS) and threshold method (TM) developed for graph coloring with the objective of solving the cellular manufacturing design problem has been presented. The results obtained are either the best one known or already of a high standard for the simulated annealing, tabu search and threshold method. Threshold method and simulate annealing were better than tabu search and sequential heuristics in terms of the inter-cell movements or the dimensions of the solutions for several examples from the literature treated. Tables 6 and 7 present, respectively, the solutions obtained by sequential heuristics (SH) or tabu search (TM), and simulated annealing (SA) or threshold method (TM) for the workshop proposed

Table 5. Computational results

E	M	C	P × M	I	D	T
Ribeiro and Meguelati, 2002	BS SH SA TS TM	2	9 × 12	0 2 0 2 0	4×6,5×6 4×6,5×6 4×6,5×6 4×6,5×6 4×6,5×6	— 0.008 1.204 1.018 1.865
Ribeiro and Pradin, 1993	BS SH SA TS TM	3	30 16	16 24 16 24 16	5×3,10×6,15×7 5×4,10×6,15×6 5×3,10×6,15×7 5×4,10×6,15×6 5×3,10×6,15×7	— 1.987 8.133 6.098 19.334
Harhalakis, Nagi, and Proth,1990	BS SH AS TS TM	5	20 20	14 15 14 15 14	5×4,4×3,4×5,3×4,4×4 6×5,4×4,4×5,3×3,3×3 5×4,4×3,4×5,3×4,4×4 6×5,4×4,4×5,3×3,3×3 5×4,4×3,4×5,3×4,4×4	— 0.985 6.124 4.769 9.897
Harhalakis, Nagi, and Proth,1990	BS SH SA TS TM	4	20 20	11 11 11 11 11	7×7,6×5,4×5,3×3 7×7,6×5,4×5,3×3 7×7,6×5,4×5,3×3 7×7,6×5,4×5,3×3 7×7,6×5,4×5,3×3	— 0.824 5.655 4.086 8.897
Ribeiro and Pradin, 1993	BS SH SA TS TM	3	20 12	0 0 0 0 0	6×4,9×5,5×3 6×4,9×5,5×3 6×4,9×5,5×3 6×4,9×5,5×3 6×4,9×5,5×3	— 1.554 3.776 2.899 7.881
Kusiak, 1987	BS SH SA TS TM	2	5 ×4	0 0 0 0 0	2×2,3×2 2×2,3×2 2×2,3×2 2×2,3×2 2×2,3×2	— 0.005 0.402 0.678 0.813

by Ribeiro and Meguelati, 2002. The manufacturing cells designed by SA or TM present 0 inter-cell moves and by SH or TS 2 inter-cell moves. The computational time to find the best solution was equal to 0.006, 0.523, 0.779 and 0.989 seconds respectively for SH, TM, SA and TM.

Table 6. Manufacturing cells obtained by SH and TS (2 inter-cell moves)

	M 1	M 2	M 4	M 6	M 8	M 11	M 3	M 5	M 7	M 9	M 10	M 12
Part 1	1	1		1	1	1						
Part 2	1	1	1	1	1							
Part 3	1		1	1	1							
Part 5		1	1		1	1					**1**	
Part 6	1	1	1		1	1						
Part 4							1		1	1	1	
Part 7								1		1	1	
Part 8							1	1	1			
Part 9	**1**						1	1		1		1

Table 7. Manufacturing cells obtained by SA and TM (0 inter-cell moves)

	M 3	M 5	M 7	M 9	M 10	M 11	M 1	M 2	M 4	M 6	M 8	M 12
Part 4	1		1	1	1							
Part 5	1	1		1	1	1						
Part 7		1			1	1						
Part 8	1	1	1									
Part 1							1	1		1	1	1
Part 2							1	1	1	1	1	
Part 3							1		1	1	1	
Part 6							1	1	1		1	1
Part 9							1	1	1		1	1

The four algorithms take very little computational time. Thus, it is sure of obtaining a solution, which is either optimal or feasible within a very reasonable computational time. Nevertheless, for large examples generated at random, the metaheuristic tabu search presented best results than simulated annealing for the same computational time.

When the computational time is free, the results obtained generally are the same or threshold method is better.

The procedure of dissimilarity indexes computation is a very important step of the method. This parameter has a hard role in the partitioning of parts, because it measures the "distances" between parts. With this in mind, research on the computation of differences between parts was recently conducted in DMB implement factory for sugar cane located in Brazil, where 3,500 parts are manufactured by 101 machines. The objective of this study is to explore the specific characteristics of the industrial real case instances in order to obtain a high standard solution.

REFERENCES

Adil, G. K., & Ghosh, J. B. (2005). Forming group technology cells incrementally using grasp. *International Journal of Advanced Manufacturing Technology*, *26*(11), 1402–1408. doi:10.1007/s00170-004-2113-8

Aho, A., Hopcroft, J., & Ullman, J. (1983). *Data structure and algorithms*. Addison Wesley.

Albadawi, Z., Bashir, H. A., & Chen, M. (2005). A mathematical approach for the formation of manufacturing cells. *International Journal on Computers and Industrial Engineering*, *48*(1), 3–21. doi:10.1016/j.cie.2004.06.008

Askin, R. G., & Chiu, K. S. (1990). A graph partitioning procedure for machine assignment and cell formation in group technology. *International Journal of Production Research*, *28*(1), 1555–1572. doi:10.1080/00207549008942812

Baroni-Urban, C., & Buser, M. W. (1976). Similarity of binary data. *Zoology (Jena, Germany)*, *25*, 251–259.

Baykasoglu, A. (2003). Capability-based distributed layout approach for virtual manufacturing cells. *International Journal of Production Research*, *41*(11), 2597–2618. doi:10.1080/0020754031000087229

Boulif, M., & Atif, K. (2006). A new branch-and-bound-enhanced genetic algorithm for the manufacturing cell formation. *International Journal on Computers and Operations Research*, *33*(8), 2219–2245. doi:10.1016/j.cor.2005.02.005

Burbidge, J. L. (1975). *The introduction of group technology*. John Wiley.

Cao, D., & Chen, M. A. (2005). A robust cell formation approach for varying product demands. *International Journal of Production Research*, *43*(8), 1587–1605. doi:10.1080/00207540412331327754

Caux, C., Bruniaux, R., & Pierreval, H. (2000). Cell formation with alternative process plans and machine capacity constraints. *International Journal of Production Economics*, *64*(1), 279–284. doi:10.1016/S0925-5273(99)00065-1

Christofides, N. (1975). *Graph theory: An algorithmic approach*. Academic Press.

Chu, C. H., & Hayya, J. C. (1991). A fuzzy clustering approach to manufacturing cell formation. Proceedings of the 1991 International Industrial Engineering Conference, (pp. 495-500). Miami, USA.

Diallo, M., Pierreval, H., & Quilliot, A. (2001). Manufacturing cells design with flexible routing capability in presence of unreliable machines. *International Journal of Production Economics*, *74*(1), 175–182. doi:10.1016/S0925-5273(01)00124-4

Dimopoulos, C., & Mort, N. (2001). A hierarchical clustering methodology based on genetic programming for the solution of cell-formation problems. *International Journal of Production Research*, *39*(1), 1–19. doi:10.1080/00207540150208835

Foulds, L. R., French, A. P., & Wilson, J. M. (2006). The sustainable cell formation problem: Manufacturing cell creation with machine modification costs. *International Journal on Computers and Operations Research*, *33*(4), 1010–1032. doi:10.1016/j.cor.2004.09.001

Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. Freeman.

Glover, F. (1989). Tabu search – Part I. *ORSA Journal on Computing, 1*, 190-206.

Glover, F. (1990). Tabu search – Part II. *ORSA Journal on Computing*, *2*, 4–32.

Gonçalves, E. V., & Tiberti, A. J. (2006). A genetic algorithm for the machine cell formation problem. *International Journal of Production Economics*, *102*(1), 1–21. doi:10.1016/j.ijpe.2004.12.029

Gonçalves, J. F., & Resende, M. G. C. (2004). An evolutionary algorithm for manufacturing cell formation. *International Journal on Computers and Industrial Engineering*, *47*(2), 247–273. doi:10.1016/j.cie.2004.07.003

Groover, M. P. (1987). *Automation production systems and computer integrated manufacturing*. Prentice Hall.

Guénoche, A. (1993). Enumeration des partitions de diamètre minimum. [in French]. *Discrete Mathematics*, *1*(11), 227–287.

Guerrero, F., Lozano, S., Smith, K. A., & Canca, D., 7 Kwok, T. (2002). Manufacturing cell formation using a new self-organizing neural network. *International Journal on Computers and Industrial Engineering*, *42*(2), 377–382. doi:10.1016/S0360-8352(02)00039-6

Harhalakis, G., Nagi, R., & Proth, J. M. (1990). An efficient heuristic in manufacturing cell formation for group technology applications. *International Journal of Production Research*, *28*(1), 185–198. doi:10.1080/00207549008942692

Hertz, A. (1991). Cosine: A new graph coloring algorithm. *Operations Research*, *10*, 411–415.

Hicks, C. (2004). A genetic algorithm tool for designing manufacturing facilities in the capital goods industry. *International Journal of Production Economics*, *90*(2), 199–211. doi:10.1016/S0925-5273(02)00467-X

Hyer, N. L., & Wemmerlöw, U. (1989). Group technology in the United States manufacturing industry. *International Journal of Production Research*, *27*(8), 1287–1304. doi:10.1080/00207548908942623

Islam, K. M., & Sarker, B. P. (2000). A similarity coefficient measure and machine-parts grouping in cellular manufacturing systems. *International Journal of Production Research*, *38*(3), 699–720. doi:10.1080/002075400189374

Jeon, G., & Leep, H. R. (2006). Forming part families by using genetic algorithm and designing machine cells under demand changes. *International Journal on Computers and Operations Research*, *33*(1), 263–283. doi:10.1016/j.cor.2005.03.033

Kirkpatrick, S., Gellat, C. D., & Vecchi, M. P. (1983). Simulated annealing. *Science*, *220*, 671–680. doi:10.1126/science.220.4598.671

Korman, S. M. (1979). The graph coloring problem In *Combinatorial optimization*. John Wiley.

Kusiak, A. (1987). The generalized group technology concept. *International Journal of Production Research*, *26*(5), 561–569. doi:10.1080/00207548708919861

Lozano, S., Canca, D., Guerrero, F., & Garcia, J. M. (2001). Machine grouping using sequence-based similarity coefficients and neural networks. *Robotics and Computer-integrated Manufacturing*, *17*(5), 399–404. doi:10.1016/S0736-5845(01)00015-1

Matula, D. W., Marble, G., & Isaacson, J. D. (1972). Graph coloring algorithms In *Graph theory and computing*. Academic Press.

McAuley, J. (1972). Machine grouping for efficient production. *Production Engineering*, *51*(2), 53–57. doi:10.1049/tpe.1972.0006

Moiser, C. T. (1989). An experiment investigating the application of clustering procedures and similarity coefficients to the group technology cell formation problem. *International Journal of Production Research*, *27*(10), 1811–1835. doi:10.1080/00207548908942656

Oliveira, S., Ribeiro, J. F. F., & Seok, S. C. (2008). A comparative study of similarity measures for manufacturing cell formation. *Journal of Manufacturing Systems*, *27*(1), 19–25. doi:10.1016/j.jmsy.2008.07.002

Oliveira, S., Ribeiro, J. F. F., & Seok, S. C. (2009). A spectral clustering algorithm for manufacturing cell formation. *International Journal on Computers and Industrial Engineering*, *57*(3), 1008–1014. doi:10.1016/j.cie.2009.04.008

Panchalavarapu, P. R., & Chankong, V. (2005). Design of cellular manufacturing systems with assembly considerations. *International Journal on Computers and Industrial Engineering*, *48*(3), 449–469. doi:10.1016/j.cie.2004.06.010

Pashkevich, A., & Kazheunikau, M. (2005). Neural network approach to trajectory synthesis for robotic manipulators. *Journal of Intelligent Manufacturing*, *16*(2), 173–187. doi:10.1007/s10845-004-5887-5

Rajagopalan, A., & Fonseca, D. J. (2005). Volume sensitivity analysis for manufacturing cells: A genetic algorithm. *Journal of Advanced Manufacturing Systems*, *4*(2), 167–183. doi:10.1142/S0219686705000667

Rajagopalan, R., & Batra, J. L. (1975). Design of cellular production systems: A graph-theoretic approach. *International Journal of Production Research*, *13*(6), 567–579. doi:10.1080/00207547508943029

Ramabhatta, V., & Nagi, R. (1998). An integrated formulation of manufacturing cell formation. *Operations Research*, *77*, 79–95.

Rath, S. K., Das, C., & Sahu, S. (1995). Graph colouring approach for cell formation problems. *Production Planning and Control*, *6*(5), 421–428. doi:10.1080/09537289508930299

Ribeiro, J. F. F. (2003). Selection and assignment of machines: A parallel approach. *Journal of Computing and Information Technology*, *11*(4), 263–270. doi:10.2498/cit.2003.04.01

Ribeiro, J. F. F. (2009). *Manufacturing cells formation based on graph theory.* Paper presented at 39th International Conference on Computers & Industrial Engineering, Paris, France.

Ribeiro, J. F. F., & Meguelati, S. (2002). Organização de um sistema de produção em células de fabricação. [in Portuguese]. *Journal Gestão e Produção*, *9*(1), 62–77.

Ribeiro, J. F. F., & Pradin, B. (1993). A methodology for cellular manufacturing design. *International Journal of Production Research*, *31*(1), 235–250. doi:10.1080/00207549308956723

Rios, M. C., Campbell, A. M., & Irani, S. A. (2002). An approach to the design of a manufacturing cell under economic considerations. *International Journal of Production Economics*, *78*(3), 223–237. doi:10.1016/S0925-5273(02)00102-0

Romesburg, C. (1984). *Cluster analysis for researchers*. Lifetime Learning Publications.

Sarker, B. P. (1996). The resemblance coefficients in group technology: a survey and comparative study of relational metrics. *International Journal on Computers and Industrial Engineering*, *30*(1), 103–116. doi:10.1016/0360-8352(95)00024-0

Sarker, B. P., & Islam, K. M. (1999). Relative performance of similarity and dissimilarity measures. *International Journal on Computers and Industrial Engineering*, *37*(7), 769–807. doi:10.1016/S0360-8352(00)00011-5

Selim, H. M. (2002). Manufacturing cell formation problem: a graph partitioning approach. *Industrial Management & Data Systems*, *102*(6), 341–352. doi:10.1108/02635570210432046

Selim, H. M., & Abdel, A. (2003). Formation of machine groups and part families: A modified SLC method and comparative study. *Integrated Manufacturing Systems*, *14*(2), 123–137. doi:10.1108/09576060310459429

Singh, N. (1993). Design of cellular manufacturing systems: An invited review. *European Journal of Operational Research*, *69*, 284–291. doi:10.1016/0377-2217(93)90016-G

Slomp, J., Chowdary, B. V., & Suresh, N. (2005). Design of virtual manufacturing cells: A mathematical programming approach. *Robotics and Computer-integrated Manufacturing*, *21*(3), 273–288. doi:10.1016/j.rcim.2004.11.001

Solimanpur, M., Vrat, P., & Shankar, R. (2004). A multi-objective genetic algorithm approach to the design of cellular manufacturing systems. *International Journal of Production Research*, *42*(7), 1419–1441. doi:10.1080/00207540310001638073

Spiliopoulos, K., & Sofianopoulou, S. (2003). Designing manufacturing cells: A staged approach and a tabu search algorithm. *International Journal of Production Research*, *41*(11), 2531–2546. doi:10.1080/0020754031000088156

Suer, G. A., Peña, Y., & Vázquez, R. (2003). Cellular design to minimize investment costs by using evolutionary programming. *International Journal on Manufacturing Technology and Management*, *5*(12), 118–132. doi:10.1504/IJMTM.2003.002536

Suer, G. A., Vásquez, R., & Peña, Y. (1999). Cellular design to minimize investment costs by using evolutionary programming. *Journal of Engineering Valuation and Cost Analysis*, *5*(1-2), 133–142.

Vin, E., De Lit, P., & Delchambre, A. (2005). A multiple-objective grouping genetic algorithm for the cell formation problem with alternative routings. *Journal of Intelligent Manufacturing*, *16*(2), 189–205. doi:10.1007/s10845-004-5888-4

Wei, J. C., & Kern, G. M. (1989). Commonality analysis: A linear cell clustering algorithm for group technology. *International Journal of Production Research*, *27*(12), 2053–2062. doi:10.1080/00207548908942674

Welsh, D. J. A., & Powell, M. B. (1967). An upper bound for the chromatic number of a graph and its application to time–tabling problems. *The Computer Journal*, *10*, 85–96. doi:10.1093/comjnl/10.1.85

Werra, D. (1990). Heuristics for graph coloring. *Computing*, *7*(Supplementum), 191–208.

Won, Y. (2000). Two–phase approach to group technology cell formation using efficient p–median formulations. *International Journal of Production Research*, *38*(7), 1601–1613. doi:10.1080/002075400188744

Xambre, A. R., & Vilarinho, P. M. (2003). A simulated annealing approach for manufacturing cell formation. *European Journal of Operational Research*, *151*(2), 434–446. doi:10.1016/S0377-2217(02)00836-6

Xu, H., & Wang, H. P. (1989). Part family formation for group technology applications based on fuzzy mathematics. *International Journal of Production Research*, *27*, 1637–1651. doi:10.1080/00207548908942644

Yin, Y., Yasuda, K., & Hu, L. (2005). Formation of manufacturing cells based on material flows. *International Journal of Advanced Manufacturing Technology*, *38*(2), 385–395.

Zhao, C., & Wu, Z. (2000). A genetic algorithm for manufacturing cell formation with multiple routes and multiples objectives. *International Journal of Production Research*, *38*(2), 385–395. doi:10.1080/002075400189473

This work was previously published in Operations Management Research and Cellular Manufacturing Systems: Innovative Methods and Approaches, edited by Vladimir Modrák and R. Sudhakara Pandian, pp. 53-67, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 93
Ethos as Enablers of Organisational Knowledge Creation

Yoshito Matsudaira
Japan Advanced Institute of Science and Technology, Japan

ABSTRACT

This chapter considers knowledge creation in relation to improvements on the production line in the manufacturing department of Nissan Motor Company and aims to clarify the substance that enables such knowledge creation. For that purpose, firstly, embodied knowledge observed in the actions of organisational members who enable knowledge creation will be clarified. By adopting an approach that adds a first, second, and third-person's viewpoint to the theory of knowledge creation, this research will attempt to define enablers of knowledge creation. Embodied knowledge, observed in the actions of organisational members who enable knowledge creation, is the continued practice of ethos (in Greek) founded in ethics and reasoning. Ethos is knowledge (intangible) assets for knowledge creating companies. Substantiated analysis classifies ethos into three categories: the individual, team and organisation. This indicates the precise actions of the organisational members in each category during the knowledge creation process and it is easier to commit further to knowledge creation activities.

INTRODUCTION

My chapter examines production improvement activities in the automobile manufacturing industry as a process of knowledge creation, and aims to clarify the factors that enable knowledge creation in a process of knowledge creation (the SECI process by Nonaka & Takeuchi).

DOI: 10.4018/978-1-4666-1945-6.ch093

For the sake of analysis, it adopts a knowledge creation theory (Nonaka & Takeuchi, 1995) to which have been added first, second and third-person viewpoints (Matsudaira & Fujinami, 2008). In this study, the statements of the interviews are analysed and interpreted through the framework of categories of socialisation, externalisation, combination, internalisation, which constitute the knowledge-creating process of the theory of knowledge creation, to which first, second, and third-person's viewpoints were added. I apply the proposed framework to the case of Oppama plant of Nissan Motor Company. By newly introducing first, second, and third-person's viewpoints, it will be possible to achieve a much better way of utilising the theory of knowledge creation within the actual situations than was ever possible. The inclusion of these viewpoints make possible a grasp, not possible using a past theory of knowledge creation, of action and the subjective aspects of organisational members who create knowledge. And from that it will be possible to elucidate the factors, which enable knowledge creation.

This chapter consists of ten sections. It begins with a review of literature on knowledge-based management, centring on the conventional treatment of the enablers of knowledge creation. Sections 3 and 4 provide theoretical reviews of knowledge-creating process. Section 5 describes the standpoints, from which interviews of the people engaged in the knowledge-creating process were conducted in order to clarify the enablers of knowledge creation. Section 6 elucidates the background of the object of the study and explanations, and Section 7 describes the research methodology. Section 8 is the main part of this chapter, aiming to show ethos as the enabler of knowledge creation through analysis and interpretation of the interviews. The discussion of possibilities for future research in Section 9 leads to the conclusion, which is last section.

LITERATURE REVIEW

In this chapter, I will try to clarify the enablers of knowledge creation, which are involved in the process of production line improvement. The problem of enablers of knowledge creation has been discussed in knowledge-based management. I would like to start by surveying the research on the topic and by showing how the enablers of knowledge creation have been treated in research. Through this I will demonstrate that this research views enablers of knowledge creation from a quite different viewpoint.

The theory of knowledge creation lists 'organisation's intention', 'autonomy', 'fluctuation and creative chaos', 'redundancy' and 'requisite variety' as enablers of knowledge creation. These are necessary when a manager is supervising the knowledge-creating process on the level of organisation.

Von Krogh et al. (2000) defines knowledge enabling as the whole set of organisational activities, which positively affect knowledge creation, stating that 'knowledge enabling includes facilitating relationships and conversations as well as sharing local knowledge across an organisation or beyond geographic and cultural borders '(p4). As the examples of knowledge enablers, they names the following: 'instilling a knowledge vision', 'managing conversations', 'mobilising knowledge activists', 'creating the right context' and 'globalising local knowledge'. These are also suggested from a viewpoint of a manager, who aids knowledge creation on organisational level.

Nonaka & Toyama (2005) present a dynamic model of a knowledge creating company and state the factors, which enable knowledge creation. Here they list 'knowledge vision', 'driving objectives', 'dialogue', 'practice', 'ba (shared context) ', 'knowledge assets' and 'environment' as enablers. These enablers also animate knowledge creation on the organisational level

The enablers of knowledge creation mentioned above are suggested from the viewpoint of a manager who aids knowledge creation of organisational members on the organisational level. Moreover, these enablers are connected with how an organisation as a whole should motivate the organisational members, who create knowledge, as well as what kind of relationships members should build between themselves and with the outside environment.

This research does not try to view the enablers of knowledge creation on a macro organisational level as did von Krogh et al., as slogans, as points to pay attention to, as actions, which managers should have in mind in order to motivate organisational members who create knowledge, or as ways to create relationships between organisational members or between members and the outside environment. Here I will define enablers on a micro level as things and ways organisational members used in order to create knowledge. I will also take into consideration subjective aspects such as sense of values of organisational members.

In order to grasp such subjective aspects as concrete actions of organisational members, which were needed for creating such knowledge, and also the sense of values, which serves as a basis for these actions, this research will use a framework, which introduces first, second, and third-person's viewpoint (Matsudaira & Fujinami, 2008) into the theory of knowledge creation described above.

Most of the enablers examined by preceding research are externally operating factors, separated from the actual subject of knowledge creation. However, the enablers in our research are intrinsic to the active subject of knowledge creation, are inseparable from it.

PEOPLE'S ABILITIES, NECESSARY FOR KNOWLEDGE CREATING PROCESS

Although I will treat these production improvement activities as a knowledge-creating process, first I would like to look into the SECI model of knowledge creation (Nonaka & Takeuchi, 1995; Nonaka, 1991, 1994). SECI model supposes that knowledge is created through a conversion process of explicit knowledge and tacit knowledge of four phases: socialisation, externalisation, combination, and internalisation (Nonaka et al., 2000, pp.9-10).

1. **Socialisation:** Socialisation is the process of converting new tacit knowledge through shared experiences. Tacit knowledge can be acquired only through shared experience, such as spending time together or living in the same environment. Socialisation typically occurs in a traditional apprenticeship, where apprentices learn the tacit knowledge needed in their craft through hands-on experience, rather than written textbooks or manuals.
2. **Externalisation:** Externalisation is the process of articulating tacit knowledge into explicit knowledge. When tacit knowledge is made explicit, knowledge is crystallised, thus allowing it to be shared by others, and it becomes the basis of new knowledge. Concept creation in new product development is an example of this conversion process.
3. **Combination:** Combination is the process of converting explicit knowledge into more complex and systematic sets of explicit knowledge. Explicit knowledge is collected from inside or outside the organisation and then combined, edited or processed to form new knowledge. The new explicit knowledge is then disseminated among the members of the organisation. The combination mode of knowledge conversion can also include the 'breakdown' of concepts. Breaking down a concept such as a corporate vision into operationalised business or product concepts also creates systematic, explicit knowledge.
4. **Internalisation:** Internalisation is the process of embodying explicit knowledge into tacit knowledge. Through internalisation, explicit knowledge created is shared throughout an organisation and converted into tacit knowledge by individuals. Internalisation is closely related to 'learning by doing'. Explicit knowledge, such as product concepts or manufacturing procedures, has to be actualised through action and practice.

Socialisation and externalisation emphasises feeling, that is, direct experience conveyed via the body, and intuition. That which is derived from feeling is reality-related raw information. The importance of real, on-the-spot experience is often emphasised because it yields information that cannot be obtained simply by reading data collected by others. Information derived from reality contains many items that cannot be converted into data. Some measurement techniques are necessary in order to generate data; however, no measurement techniques exist for many phenomena. When we consider phenomena that we cannot measure, our only recourse is to gather information with our own body.

In combination, logical thinking plays an important role. However, it is difficult to promote combination by thinking alone. For example, it is not easy to maintain motivation at the stage of planning of improvements of the production line. Emotional energy should provide a drive for thinking.

In internalisation, tacit knowledge gains results in assessments that go beyond reason, that is to say, in empirical judgments (sense of values). Excellent empirical judgments are cultivated by the repetition of high-quality experiences. A sense of values may be defined as the abstract expression of empirical knowledge related to likes and dislikes. An individual forms his or her own sense of values via experience. A sense of values provides criteria to assess what is important, and this may differ depending on the person.

TRI-DIRECTIONAL APPROACH TO ORGANISATIONAL KNOWLEDGE CREATION

In "socialisation", "externalisation", "combination" and "internalisation" – the expressions used in SECI model – the analytic viewpoint is located outside the organisation. It is a viewpoint of a transcendent being looking down at the organisation. It is as if we are looking at a surgery through a glass. My aim in this paper is to look at knowledge creation not from a viewpoint outside the organisation, but to look at it from within, from the viewpoint of an agency belonging to an organisation, i.e. the iFA members, and direct my gaze to the concrete actions and the subjective side of the actors who perform knowledge creation.

In socialisation and internalisation, when the subject of research is intuitive understanding and empirical judgments (sense of values), the question is how they should be handled. Intuition and emotion are subjective phenomena. They are difficult to handle because it is difficult to understand other individuals. The only way to proceed is to listen to the reports of the parties concerned regarding the workings of the intuition and emotion of individuals. Therefore, it is indispensable to interview the parties concerned.

This technique may be the best we have at our disposal, even though it is not easy to gain admittance to the hearts of the parties concerned. We do have the skill of sympathy. By making comparisons with our own experience, we can assess the accuracy of the intuition of others and share their enjoyment and grief. We cannot objectively describe the intuition and emotion of others from the standpoint of the observer. However, sympathy allows us to talk about the heart of others.

Moreover, circumstantial evidence is not entirely meaningless; it can provide supporting facts. The presence of subjective reports and of objective facts with consistency enhances the accuracy of subjective reports. Objective facts consist of products and services as results as well as sketches or memos that were produced at the various stages of development of the products and services. What we seek are descriptions of the internal world backed by objective facts. Explanations based on cause and effect are impossible in the case of objective facts because they cannot act as the causes of subjective reality. However, objective facts can become the results of subjective reality. Therefore, subjective reality can be envisioned by backtracking to cause and effect.

There is the subjective reality of the parties concerned, understanding via the sympathy of neighbouring persons, and supporting evidence provided by objective facts. These are thought to be subjective phenomena. As such, they are three different ways of approaching intuition and emotion. It is thought that approaching from three directions at the same time may enable research of intuitive understanding and empirical judgment (sense of values). The three methods of approach may be named as follows:

[the first-person's viewpoint] the subjective reality of the parties concerned
[the second-person's viewpoint] understanding via the sympathy of neighbouring persons
[the third-person's viewpoint] supporting evidence provided by objective facts

It is necessary to observe and analyse phenomena in an integrated manner from three different viewpoints in order to understand human beings and organisations. Varela and Shear (1999) pointed out the importance not only of a third-person viewpoint, but of a first- and second-person viewpoints, too. Essentially, subjective phenomena must be handled raw, and thus are not considered science but are treated in combination with the second-person's viewpoint and the third-person's viewpoint.

WAYS OF REPRESENTING VIEWPOINTS

Socialisation and Internalisation

We believe that we can basically organise the actuality of the creative activities that we have approached from three different directions by following the SECI model.

In the SECI model, knowledge creation is regarded as consisting of the following four phases: socialisation, externalisation, combination, and internalisation. In socialisation, we want to ascertain what the problem is by means of our five senses. In the case of quest, the quest is assigned a direction under a sense of values, that is, under the influence of having determined what is important. When this phase is examined from the three viewpoints, it is necessary to carefully observe the following points:

(1) **Socialisation**

When this phase is examined from the three viewpoints, it is necessary to carefully observe the following points:

[the first-person's viewpoint] The points that should be carefully observed are those places which explained the discovery of the issues that needed to be addressed.
[the second-person's viewpoint] If the leading role is assigned to the parties concerned, focus on the points that show how and when those people in supporting roles accepted the motivation and goals of the parties concerned.
[the third-person's viewpoint] Carefully observe the wide range of attempts and trials that were undertaken in the process of problem-finding, for example, whether attempts were made by the group to visit those places where customers live.

(2) **Internalisation**

In the case of internalisation, we organise whether and how a sense of values was obtained or verified. Clarify the differences in the sense of values before the launch of the project and after it ended. It is necessary to carefully observe the following points:

[the first-person's viewpoint] Things considered to be important (for example, customer satisfaction). Is there satisfaction with the results of the work?
[the second-person's viewpoint] Opinions about the optimal company culture and organi-

sation (human relations), the goals of the organisation, and its course of action.

[the third-person's viewpoint] The described contents of the project shareable within the organisation.

Externalisation and Combination

The interest, which comes from a position that values explicit knowledge, is externalisation and combination.

(1) **Externalisation**

In "externalisation", as well as in "socialisation", the ability that aids knowledge creation is considered to be feeling and intuition, so the hypothesis is generated by these two. The formation of a hypothesis is two-sided, containing both a selective eye and the ability for sympathy. It is necessary to form a hypothesis so that it becomes an answer to the ultimate goal of how things should be.

Imaginative power is important in concept creation and the abilities such as using keywords or visualising ideas function effectively. However, it is important not only to conceive, but also to have a selective eye that picks out only those buds which can eventually bear fruit. When this phase is examined from the three viewpoints, it is necessary to carefully observe the following points:

[the first-person's viewpoint] How to conceive the proper state of affairs.

[the second-person's viewpoint] How do the members interpret the idea, what is their reaction and the impact of the idea on them (sympathy).

[the third-person's viewpoint] The mechanisms and the technology which were actually built.

(2) **Combination**

"Combination" is a stage, in which the practical use of information, made through the combination of explicit knowledge, and the systematisation of knowledge and explicit knowledge of two or more groups on organisational level are combined. The ability to support knowledge creation in "combination" is, as in the case of "internalisation", thinking and emotion. In "combination", the actor performs theory formation. The ability to coordinate everyone when needed so, that a feeling of unity can be obtained, becomes important. It is also an important task to motivate individuals and maintain the incentive. When this phase is examined from the three viewpoints, it is necessary to observe carefully the following points:

[the first-person's viewpoint] The choice of what can be used.

[the second-person's viewpoint] Methods of advancing discussion based on that selection.

[the third-person's viewpoint] Design of the whole based on selection.

BACKGROUND

In Nissan Motor Company (Nissan), the manufacturing ability of production plants in Mexico or China has improved, and their gap with the manufacturing ability of the factories in Japan is getting smaller. There was a sense of impending crisis in Oppama plant (Kanagawa prefecture, Japan), which is one of the manufacturing factories of Nissan in Japan: if they do not raise the manufacturing ability, others will catch up. Oppama plant produces the so-called compact cars such as March and Cube, the popularly-priced cars selling for 1,400,000 to 1,700,000 yen. On the other hand, Tochigi plant (Tochigi prefecture, Japan), produces such cars as President (about 10 million yen), which is classified into "sports & speciality" class, and Skyline (sedan type, about 3,800,000 yen). Oppama plant requires its production line to be less costly than the one used for the types of cars produced in Tochigi plant.

For example, when we compare wages per hour, if we set Renault as 100%, in Japan they would constitute 120-130%, and in Mexico only 10%. If we take this concrete cost difference into account, it is better to produce Cube of the next term in Mexico, rather than in Japan. If Oppama plant does not reach a cost level, which would surpass Mexico, the next term type Cube will be produced in Mexico, it will not be possible to produce it in Oppama plant, and this would at the same time mean a serious downsising of employees. Thus, the Oppama Challenge Declaration of "Reaching a Cost Level Surpassing Mexico" was issued, and the full-scale operation began in July, 2006.

What is Oppama Challenge?

Oppama challenge is a "survival strategy for re-capturing quantity-of-production." The idea is to swiftly perform "a reform in manufacturing ability" through bench marking with low displacement cars and LCC (Low Cost Competitive) countries, as well as through intensification of collaboration between the construction site and technology, and ultimately aim for the world best factory in terms of QCT (Quality Cost Time). This reform completely eradicates doubts about factory cost competitiveness of Japan and greatly contributes to improvement in manufacturing ability of the whole of Nissan Corporation. It aims at attracting cars from all over the world by the newly attained world best manufacturing ability and freeing itself from the status of a factory for domestic use.

What is iFA?

Oppama challenge consists of six teams and one of them is iFA. iFA stands for "integrated Factory Automation". It is a combined team consisting of 18 persons assembled from various divisions of Nissan, such as the Body Division of the Manufacture-Department, the Assembly Section, the Production Division of Engineering Department and the Engineering Division. iFA means rectification (streamlining) of the flow of things and cheap automation. The effort is aimed at paying attention to the flow of things within iFA and constructing a situation, where nothing is wasted and the flow has no backwater. In order to achieve this goal, there is a need for an ingenious and cheap automation "Karakuri[1)]". If an enormous amount of money is spent on automation, there would be no cost reduction. The aim is to perform a rectification (streamlining) of the flow and achieve a situation where nothing is wasted.

The basis for iFA members' way of thinking about production is the so-called NPW (Nissan Production Way) (Nissan Motor Company, 2005), which Nissan set forth beginning from 1994. Let us briefly explain the main points of NPW.

Nissan Production Way considers a mechanism of manufacturing that transforms input to output to be the "Production System." It can be said that the profitability and competitiveness in production activity are the results of the quality of this system management performance measures and QCT.

Aim of NPW is to improve the company's "Profitability and Competitiveness" and accomplish the above, we must construct the total optimisation of "Production System", which thoroughly synchronises with the needs of our customers.

There is the concept of "Two Never Ending" in NPW.

1. Never ending synchronisation (Douki) our manufacturing with the customers
2. Never ending quests to identify problems and in place solutions

"Never ending synchronise (Douki) our manufacturing with the customers" is to purchase the following three Douki in depth in order to develop trust in the customer, bring the presence closer to the customer, and develop long-term Nissan fans in the customer by offering high-quality products and services.

- **Synchronisation of Quality:** To build in quality that our customers demand in processes
    - Never receive any defects
    - Never produce any defects
    - Never pass any defects
- **Synchronisation of Cost:** Through Elimination of Wastes that our customers would not wish to pay for
- **Synchronisation of Time:** Delivery of our products and services our customers on timeReduction of lead-time for production and development to synchronise our production with our customers as closely as possible

"Never ending quests to identify problems and in place solutions" takes the negative "problems", which tend to be concealed until nowadays, and proactively reveals them; therefore, it enables us to see these problems in a positive light, by recognising the gap between the "Want-to be condition of manufacturing" and the "Current condition", as opportunities (chances) for improvement and innovation.

RESEARCH METHOD

As a research method I used informal interviews, where items of inquiry and order are not standardised. The reason for this choice is that an interview can be performed flexibly, that adjusting to subject's interests or the flow of conversation is easy, and that a researcher, when needed, can ask the subject in detail about points he is doubtful about or matters related to a reply.

THE ANALYSIS AND THE INTERPRETATION OF INTERVIEWS

This research treats the activities of production line improvement at Nissan Oppama plant as a knowledge-creating process. The purpose of analysis and interpretation is to clarify the actions of organisational members, which enable knowledge creation. For that reason, the statements of the interviews are analysed and interpreted through the framework of categories of socialisation, externalisation, combination, internalisation, which constitute the knowledge-creating process, to which first, second, and third-person's viewpoints were added.

Table 1. Outline of interview

Date	Length of time	Place	Interviewee
16th October 2007	About 3 hours	Nissan Oppama plant (Kanagawa prefecture)	iFA sub-leader four iFA members
4th December 2007	About 3 hours	Nissan Oppama plant	iFA leader iFA sub-leader four iFA members
4th April 2008	About 3 hours	Nissan Oppama plant	iFA member
25th May 2008	About 2 hours	Tokyo	iFA member
22nd September 2008	About 2 hours	Nissan Technical Centre (Kanagawa prefecture)	iFA leader
26th December 2008	About 2 hours	Nissan Oppama plant	iFA member
1st July 2009	About 2 hours	Nissan Technical Centre	iFA member

Enablers of Knowledge Creation

I will look into the interviews and examine the enablers of knowledge creation for each of the processes: socialisation, externalisation, combination and internalisation. The classification and the interpretation of the interviews are summarised in Table 2. (Each of the items at the beginning of the interpretation is a selected quote of part of a remark, made during the interviews).

Ethos as an Enabler of Knowledge Creation

From the classification and the interpretation of Table 2, from the knowledge-creating processes

Table 2. Enablers of knowledge creation (The number of each item corresponds with the classification of Table 3)

	Socialisation
First-person's viewpoint **Viewpoint:** **How was the problem, which should be tackled, discovered?**	**(1) "Taking the losses from the lower process to the upper process"** Having steadfast plans and ideals leads to problem discovery. When you are going around, looking at the production site, observing the activities of the workers, being conscious of the ideal of bringing the losses from the lower process to the upper, having in mind the question: "What could the next improvement be?", you always find some actions that are out of place. **(2) "Looking at other shops"** Directing interest not to the shops (divisions) you initially started in, but to other shops, leads to problem discovery. From such remarks as "I have been in Oppama plant for a long time, but have never seen other shops" there is a shift of subject of interest to "Let's all of us members go and see all the other shops in the factory". After going around and seeing other shops anew, there is much astonishment and many discoveries of usable points: points, which then become the raw material for problem discovery.
Second-person's view point **Viewpoint:** **How did the surrounding members accept the feelings and aims voiced by the parties concerned?**	**(3) "It all started from trying to find out, what iFA is there for"** The members of iFA did not start with actual plans or ideas to make things cheaper, or with examining concretely the points for improvement, such as the reduction of how many workers can be expected in the production process. First and foremost they thoroughly questioned the reason for being there, what they had gathered there for, and started with coming to a common denominator in terms of the way of thinking about production.
Third-person's viewpoint **Viewpoint:** **The attempts made in the process of problem-finding.**	**(4) "Sharing opinions about Douki production and bringing together ways of thinking."** The members were brought together as iFA and given two to three weeks time to conduct discussions in order to share their philosophies, create views and ways of thinking about production as a single body, iFA. Exactly because they were given this time and opportunity, because they could share their philosophies, they could progress in the same direction even when they were performing improvement related activities of quite harsh nature, "as if we were trying to squeeze water out of already dry rag" (words of one of the members). Because an attempt was made to provide them with this time and opportunity, they could share their ways of thinking about NPW, which is the ideal they were aiming at, and perform the actual activities.
	Externalisation
First-person's viewpoint **Viewpoint:** **How to conceive the optimal state of affairs?**	**(5) "Rectification of the flow of things", "ingenious automation called 'Karakuri'"** Slogans of reaching "a situation where there is no waste" and "cheap automation" are definitely pointing at the optimal state of the production line, at the way it should be. But these phrases are only expressing the goal. Judging these as not really touching on the essence of the iFA, the leader made the ideals more sophisticated, changing the "situation where there is no waste" to "rectification (streamlining) of the flow of things" and from "cheap automation" to "ingenious automation called 'Karakuri'". This change not only indicates the goal to the members, but forms an image of what they concretely need to do. **(6) " The model is the Nissan Production Way"** When asked about the ideas they have about the production line, about the basis for the direction of their efforts, all members answered: NPW. There is an ethical goodness called NPW and by having it as a basis, externalisation in the knowledge creation process, aiming at the improvement of the production line, is promoted.

continued on following page

Table 2. Continued

Second-person's view point **Viewpoint:** **How do the members interpret the ideas, what is their reaction to them and the impact of the ideas on them?**	**(7) " NPW is the foundation for the hypothesis"** A basis is necessary to formulate a hypothesis. The basis for the hypothesis, aimed at improvement of the production line, is NPW. A hypothesis is built through checking it against NPW. NPW is a sort of a resource, judged to be good in the Production Division. The image of rectification (streamlining) of the production line is drawn according to it. In externalisation the members conceive the way of thinking of NPW as an ethical goodness or resource. When they waver in their decisions, when they have to decide their priorities, it is NPW they turn to. This ethical goodness leads people of production division of Nissan to actions, as if saying: "This is what you should do". **(8) "Everyone understood when they heard the words 'moving the waste to the upper process' "** At first, faced with the problem of improving productivity, members wrestled with the choice between moving the waste either to the upper or to the lower process, each relying on their own keywords such as "getting rid of stocks" or "Douki", and faced the problems of improvement without a clear understanding of the reasons for their actions. However, thanks to the leader's concept, which used a metaphor of river (upper process in Japanese is called 'upper stream') and grasped the essence of the problem, namely "driving waste upstream", members though: "This is it. Now I get it. We just need to drive the waste up". Due to it, they noticed that the essence of the improvement related actions they were performing unconsciously was, in fact, driving the waste to the upper process.
Third-person's viewpoint **Viewpoint:** **Mechanisms which were actually built.**	**(9)Shishi Odoshi** 'Shishi Odoshi' is a type of transfer equipment using an AGV (Automatic Guided Vehicle: an unmanned cart, which carries parts to the required places in a factory) which carries boxes filled with picked parts and pours them onto a lane. Members ingenuously made a mechanism, which is cheap to produce and which easily moves the box filled with parts brought by AGV onto the rail. By analogy with up-down movement of a bamboo pipe, which inclines under the weight of water and is often set up in Japanese gardens, they called it Shishi Odoshi.
	Combination
First-person's viewpoint **Viewpoint:** **How did the members conceive the choice of what can be used?**	**(10) "Thinking whether it can be done in a simpler way"** When you are trying to improve the whole production line, you tend to think of large-scale, complicated mechanisms. Members tried to think of how they could simplify it instead. Furthermore, they put out their antennas to receive information from the outside world in the hope of gaining cues for simplifying there.
Second-person's view point **Viewpoint:** **How did the members discuss that selection?**	**(11) "Break the walls and do it together"** All iFA members came from different divisions. If they followed the way of thinking of their old divisions, there could be no teamwork. On the path to improvement iFA sets no restrictions on exchange of necessary things and able personnel. This breaks the walls members' old divisions set up in terms of ways of thinking and behavioural patterns and leads them to work on the improvements and tackle problems together as a single unit: iFA. Combination is promoted by the fact that people of various divisions participate and that explicit knowledge is being added.
Third-person's viewpoint **Viewpoint:** **Design of the whole based on the selection**	**(12) "Came to see the production line as a whole, and not in parts"** In case of improvement of a production process like that of iFA, it is vital to have a vision that covers the whole, from the beginning to the end, and not just parts. This way you cease to view problems separately. And when you see it as a whole, you can choose the sections, which should be given priority. While you are going from the sections to the whole and back in amoeboid-like movement, tacit knowledge and connection between people expands, helping to expand also the scope of the improvement.
	Internalisation
First-person's viewpoint **Viewpoint:** **Things members learned to consider as important through the activities and evaluation of work done**	**(13) "Looking at improvements in terms of the whole picture"** One of the points members think to be important is to grasp the whole picture. Being able to see the whole picture of the improvement means that the members managed to understand the improvement as a whole through the processes of socialisation, externalisation and combination and truly feel the importance of grasping the whole picture. **(14) "There is no such thing as 100% perfection"** An improvement is a continuous effort and there is no such thing as a final shape to it. The second you consider any state as final you lose any opportunities to make the production line better. Many of the members said in almost the same words: "We are still continuing", "We are not finished with improvements yet". As long as the production activities continue, the improvement goes on. These are the values of iFA.

continued on following page

Table 2. Continued

Second-person's view point **Viewpoint:** **Opinions related to corporate culture and the optimal state of the organisation, goals of the organisation, and its course of action**	**(15) "Every one of us has his own vision of what he wants to do with the plant"** All the members have their own vision of what they want to do with the production plant. This means that each of the members has a different sense of values. And this serves as the source of knowledge creation. Through the struggle between each member's respective values, organisational members cross their boundaries, dialectically producing new values. Values are internalised by the individual, driving the problem discovery in socialisation further. **(16) "We want to create the best manufacturing site in the world"** Through their belonging to iFA the members began to feel themselves as a part of the movement to "create the best manufacturing site in the world", which is the goal of the organisation. The direction of this movement is the idea that a manufacturing site has to be cheap, not wasteful and have all the necessities at hand. They came to believe in these values through their activities in iFA.
Third-person's viewpoint **Viewpoint:** **The described contents of the project shareable within the organisation**	**(17) "We'd like to find ways to tell others about our activities"** When the reality of a project like iFA is not institutionalised within the organisation and is not established as corporate culture, it is difficult to explain and describe such a reality objectively. Therefore, one of the iFA members wrote a novel, describing the activities of iFA. Moreover, with cooperation of people who participated in the project and those who did not, the activities were summarised in a story-like narrative called "Nissan Inside Story", thus providing the information not only for the people who are connected with improvements of this particular Nissan's factory, but so that it is possible to share it with all those working in Nissan.

of socialisation, externalisation, combination, and internalisation it is possible to detect the actions repeatedly performed by organisational members as factors, which enable knowledge creation. For example, in the knowledge-creating process of socialisation we can see an action repeatedly done for problem discovery, which consists of getting interested in other divisions, which members had never been a part of, and going around them for inspection. The actions repeatedly done by organisational members in each of the four phases of knowledge-creating process – socialisation, externalisation, combination and internalisation – are the intangible knowledge assets, which enable knowledge-creating process. These knowledge assets enable organisational knowledge creation through habits of organisational members.

What does this point to? It points to a certain ethos, embodied in organisational members. Ethos gradually becomes embodied through the interaction between the organisational members during the activities related to knowledge-creating process.

The word "ethics" comes from Latin word "ethica", which goes back to Greek word "ethika". Ethos originally meant a place one is accustomed to live in, and at a level of community means customs and traditions. An explanation of ethics going back to the Greek origin of the word can be found in Aristotle's Nicomachean Ethics (Ross, 1928; Ross et al., 1998; Broadie & Rowe, 2002). What needs a careful consideration is that things, which are called "ethical (ethikos)"are not the customs one acquires naturally, but those that become a routine through careful and intentional moulding (Sekine, 2001). Ethos is the habits intentionally embodied by the organisational members themselves during knowledge creation in order to implement the ethical goodness of NPW.

Classification of Ethos

If ethos, which is clarified knowledge assets, is put together, confusion arises in its accumulation and utilisation. The reason for it is that ethos is embodied or embedded within different layers: individuals, the team and the factory (organisation). In order to clarify this distinction, we introduced first, second, and third-person's viewpoints and analysed knowledge-creating process. In Table 3 we have classified ethos discussed above from three points of view: that of the individuals, the team and the factory.

Table 3. Classification of ethos by first, second, and third-person's viewpoint (The number of each item corresponds with the classification of Table 2)

	First-person's viewpoint
Ethos embodied in individuals	(1) Keeping in mind ideals and aims when looking at the site (2) Having the interest in other divisions, posts and production stages (5) Reflecting on the self in actions (6) Having ethical goodness as the norm for thoughts and actions (10) Not complicate matters (13) Sharing the philosophies (points of view, ways of thinking), and grasping the whole picture (14) Generate a sense of values through practice
	Second-person's viewpoint
Ethos embedded in the team	(3) Groping for identity, questioning raison d'etre and sharing the way of thinking about production (7) Having a basis for hypothesis creation (8) Clarifying concepts by using metaphors and analogies, clarifying unconscious acts (11) Not heeding the borders between divisions in work, thus combining different explicit knowledge (15) Building a new sense of values through synthesis of different values (16) Departing from practice when defining the course of action
	Third-person's viewpoint
Ethos embedded in the plant (organisation)	(4) Providing time and opportunity to share philosophies between members (9) Concretising, using analogies as hints (12) Conceiving plans by going from the parts to the whole and back, expanding explicit knowledge in amoeba-like movement (17) After finishing the project, leaving behind the concrete contents of activities as explicit knowledge in a story-like form

FUTURE RESEARCH

Firstly, the object of this research was a production division of an automotive manufacturer, and if we extend our scope to research and development divisions, it is possible to enrich the knowledge of this research even further, possibly discovering other types of ethos which enables knowledge creation.

Secondly, ethos, which is a high-quality tacit knowledge, as a source of knowing, created through specific time, space and human relations, is inseparably connected to a certain context, which in case of this research are the activities aimed at improving the production line by iFA in Oppama challenge. It is necessary to conduct a comparison of ethos described in this chapter with the ethos of the production line when not involved in the project we described, to see what is same and what is different.

CONCLUSION

The understanding based on the ethos, which enables knowledge creation, illuminates the individual and concrete human reality, such as what organisational members do, what sense of values they have, how they grapple with problems as an organisation etc. As a counter position to the two major strategy theories, which constitute the mainstream of present research – the positioning school (Porter, 1980) and the resource-based view of the firm (Wernerfelt, 1984; Barney, 1986, 1991, 2001) – which represent the logical analysis strategy, here I laid stress on the rehabilitation of people's wisdom, practice and action. Ethos is not just a habit, it is a habit with has ethics embodied or embedded within it. In the context of the material of this research ethos is a habit that makes possible to implement NPW – the goodness, which has been accepted and shared within the production division of Nissan. Knowledge creation is made possible through repeated spinning of SECI processes. And what has been thus revealed is an ever-repeating habit, in other words – ethos. Ethos is an ethical goodness, NPW, which has been embodied and embedded.

Next, I would like to give some suggestions to those actually working with ethos.

In this research, I have tried to elucidate a part of ethos, which lurks in the depths of human ability to create anew, ordaining knowledge creation from behind the scenes. I have shown that ethos enabled knowledge creation in terms of improvement of the production line, which made it possible to lower the cost of production to the level of Mexico, which was the goal of Oppama challenge, and also managed to secure the production of the new model of Cube for Japan and the North America. Ethos, while efficiently employing the knowledge producing tradition, stemming from accumulation of experience and inherited from generation to generation in a business organisation, is a concept that provides key to sustainable growth. The reason for this is that the continuous innovation through knowledge creation is indispensable for sustainable growth, and that ethos, while being the output produced by the knowledge-creating process, at the same time becomes the input that leads to the next knowledge-creating process.

Ethos is the intangible knowledge assets, which maintain knowledge creation. Since we introduced first, second, and third-person's viewpoints in this research, we have elucidated and categorised ethos, which enables knowledge creation, from three points of view - that of the individuals, the team and the factory (organisation). For the knowledge workers, who practice knowledge creation in a business organisation, this classification becomes an accumulation and utilisation of knowledge assets, an input for next knowledge creation and proves helpful for maintaining knowledge-creating process.

Ethos, which enables knowledge-creating process, can be shown in a figure as follows.

Let us explain the figure. The activities of improvement of the production line, which aim at implementation of streamlining of the flow - the ideal or proper condition - these activities through their process produce ethos. It is not that knowledge-creating activities, which aim at implementation of streamlining of the flow of things, become a cause, resulting in inventions and reforms, which allowed the streamlining to take place. The process between inventions and reforms and the reality gave birth to ethos of various kinds. And NPW functions as the reference of frame that gave birth to that ethos, providing the basis for it. Through actual practice of ethos, as a result of it, the various inventions and reforms, which implemented streamlining of the flow of things, could be made.

Ethos produces as its output the knowledge of the inventions and reforms, related to the production line, knowledge, which members wanted to implement. Moreover, since ethos also serves as the input of knowledge creation, it also plays its role in spinning SECI processes at high speed, increasing the quality and the quantity of knowledge, making a spiral repetition possible. Ethos,

Figure 1.

which enables knowledge creation, is embodied by organisational members and is embedded by the team and the organisation.

Finally I would like to describe the managerial implication of this research.

In corporate management, knowledge assets are the intangible assets, which, unlike physical assets, are implicit and are in a state of dynamic transformation. For a firm they are an essential resource indispensable for value production.

The knowledge assets contain not only the knowledge made explicit for example by patents, but also the implicit knowledge needed to produce it. In other words, the former is the explicit knowledge, which can be indicated formally, such as concrete improvements and designs of a production line, whereas the latter is the knowledge of ethos, which produces such improvements and designs in the first place. Since the former can be easily indexed with firm's values, it tends to be considered as knowledge assets, constituting the source of firm's competitive advantage. However, ethos, which is continuously keeping the knowledge-creating process alive, is what is truly attractive for corporate management or an organisation, although, being one of the types of implicit knowledge, it is difficult to index. What was indexed, was indexed as a result of the justified past, and in order to assess its value for the future one has to look into what is hiding behind it. And what is behind it is nothing other than ethos. It is ethos that is always functioning as the driving force for the next knowledge-creating process. In this sense, this research revealed numerous ethoses constituting intangible knowledge assets in the production department of Nissan Motor Company. They play the role of continuing and promoting the creation of knowledge, which serves as the source of sustainable competitive advantage for a firm.

Ethos, which is a knowledge asset, is by no means unchangeable. It dynamically changes, as do all intangible assets. Since the next knowledge-creating process is always set in a new context with a new locality of the problem, maybe not all the contents of ethos, but some are changed or created anew to meet the new requirements. We should not forget that in management, ethos has to be adjusted according to the context. People related to knowledge-based management should fully understand that every recreation of ethos is a new knowledge-creation. And that this point of view is vital for a continuous promotion of knowledge-creating process.

I believe that this research was successful in its role of showing the indispensability of ethos - the new concept of knowledge assets, which enables knowledge creation - for future knowledge-based management.

REFERENCES

Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science*, *42*, 1231–1241. doi:10.1287/mnsc.32.10.1231

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*, 99–120. doi:10.1177/014920639101700108

Barney, J. B. (2001). *Gaining and sustaining competitive advantage* (2nd ed.). New York: Prentice-Hall.

Broadie, S., & Rowe, C. (Eds.). (2002). *Aristotle's Nicomachean Ethics*. New York: Oxford University Press.

Drucker, P. F. (1993). *Post-capitalist society*. Oxford: Butterworth Heinemann.

Hall, T. E. (1990). *The silent language*. New York: Doubleday, Anchor Press.

Matsudaira, Y., & Fujinami, T. (2008). Understanding organisational knowledge creation from a perspective of embodied knowledge. *International Journal of Knowledge and Systems Science*, *5*(2), 19–27.

Nelson, R. R. (1991). Why do firms differ, and how does it matter? *Strategic Management Journal*, *12*, 61–74. doi:10.1002/smj.4250121006

Nissan Motor Company NPW suishin bu (Ed.). (2005). *Jissen Nissan seisan hoshiki: Nissan seisan hoshiki key word 25* [Nissan Production Way]. Tokyo: Nikkan Kogyo Shimbunsha.

Nonaka, I. (1990). Redundant, overlapping organizations: A Japanese approach to managing the innovation process. *California Management Review*, *32*(3), 27–38.

Nonaka, I. (1991). The knowledge-creation company. *Harvard Business Review*, *69*(6), 96–104.

Nonaka, I. (1994). Dynamic theory of organizational knowledge creation. *Organization Science*, *5*, 14–37. doi:10.1287/orsc.5.1.14

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. New York: Oxford University Press.

Nonaka, I., & Toyama, R. (2003). The knowledge-creating theory revisited: Knowledge creation as a synthesizing process. *Knowledge Management Research & Practice*, *1*, 2–10. doi:10.1057/palgrave.kmrp.8500001

Nonaka, I., & Toyama, R. (2005). The theory of the knowledge-creating firm: Subjectivity, objectivity and synthesis. *Industrial and Corporate Change*, *14*, 419–436. doi:10.1093/icc/dth058

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, ba and leadership: A unified model of dynamic knowledge creation. *Long Range Planning*, *33*, 5–34. doi:10.1016/S0024-6301(99)00115-6

Porter, M. (1980). *Competitive Strategy*. New York: Free Press.

Quinn, J. B. (1992). *Intelligent enterprise: A knowledge and service based paradigm for industry*. New York: The Free Press.

Ross, W. D. (Ed.). (1928). *Aristotle's Nicomachean Ethics*. Oxford: Oxford University Press.

Ross, W. D., Ackrill, J. L., & Urmson, J. O. (1998). *Aristotle's The Nicomachean Ethics*. New York: Oxford University Press.

Sekine, S. (2001). *Rinri siso no genryu* [Ethical ideas]. Tokyo: Hoso Daigaku Shinkokai.

Toffler, A. (1990). *Powershift: Knowledge, wealth and violence at the edge of the 21st century*. New York: Bantam Books.

Varela, F. J., & Shear, J. (1999). First-person methodologies: What, why, how? *Journal of Consciousness Studies*, *6*(2-3), 1–14.

von Krogh, G., Ichijo, K., & Nonaka, I. (2000). *Enabling knowledge creation*. New York: Oxford University Press.

Wernerfelt, B. (1984). A Resourced-based view of the firm. *Strategic Management Journal*, *5*, 171–180. doi:10.1002/smj.4250050207

ADDITIONAL READING

Barnard, C. I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.

Drucker, P. F. (1969). *The age of discontinuity*. New York: Harper and Row.

Fiyvbjerg, B. (2001). *Making social science matter: Why social science fails and how it can succeed again*. Cambridge: Cambridge University Press.

Fiyvbjerg, B. (2006). Making organization research matter. In S. Clegg, C. Hardy, T. Lawrence & W. Nord (Eds.), *The SAGE handbook of organization studies* (2nd ed., pp. 370-387). Thousand Oaks, CA: SAGE publications.

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*, 109–122.

Hamel, G., & Prahalad, C. K. (1994). *Competing for the future*. Boston: Harvard Business School Press.

Hutchins, E. (1990). The technology of team navigation In Galegher, J., Kraut, R. E., & Egido, C. (Eds.), *Intellectual teamwork: Social and technical bases of cooperative work* (pp. 191–220). Hillsdale, NJ: Lawrence Erlbaum.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT press.

Ichijo, K., & Nonaka, I. (2007). *Introduction: Knowledge as competitive advantage in the age of increasing globalization. Knowledge creation and management: New challenges for managers*. Oxford: Oxford University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Nonaka, I., Konno, N., & Toyama, R. (2001). Emergence of 'ba': A conceptual framework for the continuous and self-transcending process of knowledge creation In Nonaka, I., & Nishiguchi, T. (Eds.), *Knowledge emergence: Social, technical, and evolutionary dimensions of knowledge creation* (pp. 13–29). New York: Oxford University Press.

Nonaka, I., Toyama, R., & Hirata, T. (2008). *Managing flow*. New York: Palgrave Macmillan. doi:10.1057/9780230583702

Nonaka, I., Toyama, R., & Nagata, A. (2000). A firm as a knowledge-creating entity: A new perspective on the theory of the firm. *Industrial and Corporate Change*, *9*(1), 1–20. doi:10.1093/icc/9.1.1

Polanyi, M. (1958). *Personal knowledge: Toward a post-critical philosophy*. Chicago, IL: University of Chicago Press.

Polanyi, M. (1966). *The tacit dimension*. New York: Doubleday.

Porter, M. (1980). *Competitive advantage*. New York: Free Press.

Prahalad, C. K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, *68*(3), 79–91.

Rumelt, R. P. (1974). *Strategy, structure, and economic performance*. Cambridge, MA: Harvard University Press.

Simon, H. A. (1945). *Administrative behaviour*. New York: Macmillan.

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. New York: Cambridge University Press.

Teece, D. J. (1982). Toward an economic theory of the multiproduct firm. *Journal of Economic Behavior & Organization*, *3*, 39–63. doi:10.1016/0167-2681(82)90003-8

Teece, D. J. (1998). Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. *California Management Review*, *40*(3), 55–79.

Teece, D. J. (2000). *Managing intellectual capital*. New York: Oxford University Press.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, *28*, 1319–1350. doi:10.1002/smj.640

Toffler, A. (1980). *The third wave*. New York: Morrow.

Urmson, J. O. (1988). *Aristotle's Ethics*. Oxford: Blackwell.

Weber, M. (1947). *Wirtschaft und Gesellschaft Grunderiss der Sozialokonomik 3*. Tuebingen.

ENDNOTE

"Karakuri" is a type of Japanese traditional craftsmanship, which was invented in the 1770s. Its famous example is a doll, which carried tea. The distinguishing characteristic of Karakuri is that mechanisms do not rely on any dynamic force like electricity or batteries. Bearing the mechanism of Karakuri in mind, the members are working to implement cheap automation at the production site.

This work was previously published in Managing Knowledge Assets and Business Value Creation in Organizations: Measures and Dynamics, edited by Giovanni Schiuma, pp. 50-66, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 94
Engineering Design as Research

Timothy L.J. Ferris
Defence and Systems Institute, University of South Australia, Australia

ABSTRACT

Research is defined as an activity that creates new knowledge. This is often misunderstood in the engineering community as necessarily requiring a scientific contribution that advances the theory of some matter related to engineering materials or processes. Consequently, typical engineering research projects investigate physical phenomena thought likely to be interesting in potential applications or to describe the characteristics of processes used in engineering work. The results of such projects provide a fragmented, abstracted view of the phenomena investigated, which is difficult to use in engineering decision making related to contextualised situations. This chapter shows how the actual design of engineered artefacts is research because it provides knowledge of the impact of the integration of various elements of existing knowledge, which demonstrates the properties of the designs achieved through the design work and leads to discovery of solutions to the various challenges of integration discovered through the project which attempts to achieve the integration.

INTRODUCTION

Research is defined as an activity which develops new knowledge (Department of Education Training and Youth Affairs, 2001). A major challenge we have in understanding the nature of research is our preconceptions concerning the nature of knowledge, since the creation of knowledge is the objective of all research activities. In this chapter we introduce the particularly important issue of engineering activity and the act of design itself as research, that is, as methods of creating new knowledge. This is distinct from performing investigations about the engineering and design processes as an observer of those processes with the purpose of better understanding the processes. The idea of engineering activity and design being a research methodology requires explanation because it is quite different to the usual interpretation of research in engineering.

DOI: 10.4018/978-1-4666-1945-6.ch094

It is usually expected that the knowledge that is developed in processes to which we assign the name 'research' is objective and generalizable to a wide range of problems. This understanding is derived from the idea of research as implemented in the natural sciences and fields inspired by their methods, in which the goal is to find out knowledge about the phenomenon or situation which is the subject of the work, and the work is regarded as completed when a sufficiently good theory or explanation of the situation is determined.

The idea of research concerning the discovery of generalizable knowledge about the subject matter of the investigation is prevalent in engineering related research conducted in universities, by both faculty and graduate students, and in the regulations for research degrees in engineering and in the engineering research journals published in recent decades. The engineering research journals of the 1950's, particularly prior to the changes in the US education and research system that followed Sputnik, reflected a different vision of the nature of engineering research and its relation to the achievement of engineering design. The latter will be discussed later in this chapter.

The idea of generalizability of knowledge relates to the manner in which it can be applied to situations other than the narrow set of situations in which the observations were made (Lee & Baskerville, 2003). Generalizability can be established in several ways and can have different meanings dependant on the manner in which it is achieved. The natural sciences use statistical analysis of observations to determine the confidence that can be established that the null hypothesis is not wrongly rejected, and so seek to establish results as being universally applicable across all analogous cases. Fields which accept interpretivist views of knowledge may be satisfied with developing knowledge that is true within the bounds of the situation of observation, without demanding that the knowledge obtained be demonstrated for a broader range of situations (Lee & Baskerville, 2003). In general, the notion of generalizability demands the assumption that the matter of observation is uniform across the range of the generalization so that the observed sample represents a wider space (Lee & Baskerville, 2003).

When we look for objective, generalisable knowledge about subject matter we narrow our view of the nature of knowledge through the Positivist lens to concerning knowledge about external things. This is not the only view of knowledge which exists, and unnecessarily constrains research since it excludes the development of knowledge of other kinds.

KNOWLEDGE

We present now a discussion of the nature of knowledge which is necessary to develop a top-down approach to the evaluation of methods of research. We review some recent distinctions in the description of knowledge.

"Know that" is a formulation used to describe declarative knowledge, following Gilbert Ryle's distinction between "knowing that" and "knowing how" (Ryle, 1948). This distinction has been noted as significant in engineering, where both kinds of knowledge are required (Bucciarelli, 2003). Declarative knowledge is a kind that can be articulated in words or mathematics that represent ideas concerning the nature and relations of things. As such, declarative knowledge is a form of knowledge which it is fairly easy to teach and learn because it is possible to reduce the teaching or learning to recitation of the representation of the knowledge. The recitative character of declarative knowledge is discussed by Biggs (Biggs, 1999). Declarative knowledge is any knowledge which can be formulated in statements which convey the knowledge itself, whether that knowledge is generalizable across a wide range of cases or of limited applicability.

"Know how" is Ryle's formulation to describe the capacity to perform a function. This capacity is quite distinct from the capacity to describe the area of knowledge related to the function or the relevant theory. Biggs (1999) describes this kind of knowledge as "functional". Functional knowledge emphasizes the person's capacity to perform a function, not to articulate description of what is known. Functional knowledge does not preclude the ability to articulate what is known, but emphasizes ability to perform an act. The distinctive feature of functional knowledge is the knower's orientation to doing a function rather than describing something. Should someone articulate functional knowledge into a description it would be possible for another person to learn the description as declarative knowledge, which that person could recite, but not be able to perform appropriately.

A third kind of knowledge is named "knowing" in Nissen (2006), and by Biggs (1999) as "procedural knowledge", or "skill". The emphasis in this kind is on ability to choose and perform some action in an appropriate and effective manner. Nissen says that knowing how to ride a bicycle is demonstrated by mounting and actually riding a bicycle. The only test of "knowing" is a practical test in which the candidate must perform the action. Ability to articulate anything about the matter of action or a theory about the action is irrelevant to "knowing". "Knowing" contrasts with "know how", where the emphasis is on ability to perform a function, because "knowing" is usually associated with skills such as ability to discern when a particular action is apposite. The difference between "know how" to ride a bicycle and "knowing" how to ride a bicycle is that in the "know how" formulation the knower is able to perform the set-piece tasks associated with riding a bicycle, but may not yet be ready to integrate all the tasks into safe riding of a bicycle on a roadway with traffic. The "knowing" formulation refers to the more advanced achievement of being able to integrate all the set-piece tasks into effective and safe cycling in a normal situation.

The three kinds of knowledge span from the ability to abstract, describe and theorize about subject matter, through applied knowledge which enables performance of functions, to practical ability to perform whole and complex activities. These kinds of knowledge form identifiable points in the characterization of knowledge. Any particular knowledge held by a particular person will have its own combination of abstract theorizing and ability to act, and so could be described as belonging at a particular point along a continuum. Identification of the three kinds of knowledge is useful for theorizing about the nature of knowledge, and research.

Research is an activity which creates new knowledge, which could be of any kind. The new knowledge may be in an abstracted, declarative, form, or may be embedded in the capacity of a person to act appropriately as either functional or procedural knowledge. In describing research as means to produce new knowledge, it is important that the new knowledge be new to everyone. However, the knowledge advance may be of any size, including very small, and may be situationally constrained.

RESEARCH METHODS

At least some academic research degree regulations in engineering demand that the research work not apply "the normal engineering method"[1]. These regulations presume that there is clear agreement as to what constitutes a normal engineering method. There are two grounds for scepticism that there is a clearly agreed engineering method. These grounds are that:

1. There is very little teaching of method in engineering programs beyond quite rudimentary descriptions of things such as the project life-cycle. That is, descriptions of the engineering method omit detailed explanation of how the creative work of engineering

which synthesises solutions to problems is actually done. This knowledge is the purview of Design Studies, which itself reflects diversity of views about what constitutes the design process (Dorst, 2008; Gregor & Jones, 2007; March & Smith, 1995; Walls, Widmeyer, & El-Sawy, 1992).

2. Engineers, in general, discuss methodology very little. The engineering culture emphasises investigation of the technologies and principles which are applied in the technologies but says little about either how the principles are discovered or applied to achieve engineering outcomes. The measure of a good engineer is one who can successfully achieve challenging objectives (Hopkins, 1954).

Engineers tend to be a pragmatic lot who get on and do the work without extended discussions of the methods and principles upon which the work is based. Individual engineers develop a good pragmatic understanding of methods applicable to the kind of work that they do, largely through experience, including observation of the effectiveness of methods which they have attempted to use. This is an approach to knowledge consistent with the approach discussed above under the name "knowing".

The words from the research degree regulations "not apply the normal engineering method" are often interpreted to mean that design cannot be research because design is considered to be a normal engineering activity and therefore excluded. Design is a significant part of engineering work and the design of almost any product, from the simplest to the most complex, is the result of an engineering design activity.

We need to explore this issue. Design concerns development of situated knowledge. The outcome of design is knowledge which is specific to the particular need addressed by the project. The situated knowledge developed by design is specific to the particular solution which is offered. The knowledge that we developed through the design process is specific to the choice of technology and the system architecture selected and provides "proof by construction" that the design idea is valid and effective (Hevner, March, Park, & Ram, 2004). The action of design always, but specifically including the case of design of novel or innovative products, demands rigor in the use of the underlying principles to posit an appropriate design and in the evaluation of the results achieved (Hevner, et al., 2004).

To posit design as a research methodology challenges the tacit idea, held particularly strongly in the natural sciences because of their Positivist conceptualisation of knowledge and therefore significantly influencing many engineers, that knowledge created through research should be generalizable to all cases (Lee & Baskerville, 2003). The knowledge developed in a particular design project is specific to the particular design implemented through the project and the context in which, and for which, it is tested. That is, a design project produces distinctly situated knowledge of what is apposite in constrained circumstances.

Various authors have written about research methodologies which are suitable for use, or which are commonly used, in particular disciplines. These works are often designed as texts for research methods courses. Such courses are usually designed to provide students with instruction in how to conduct activities in the discipline in which they are studying a research degree. Research methods courses which teach one, or a very few research methods, leave students with the impression that proper research must be done using one of the methods described in the course.

Engineering should not be approached as a field which permits only a few particular research methods. This is because the core concern of engineering is the development of suitable solutions to address needs. Therefore methods of research in engineering need to include methods which can explore the scientific phenomena which may be applied in a solution, but also need to be inclusive of research methods which can explore the nature of the need to be addressed and the effectiveness of any proposed solution to be a solution to the need which has been presented, and could explore the methods for doing engineering work (Ferris, 2009). The evidence provided by the content of mainstream engineering research journals shows that much of the engineering research publication since the 1960s has been constrained to discovery about scientific principles which may be useful in future engineering projects. This is evidenced by the focus of journals being mainly centred on technology areas, with only a few, less well-known, journals addressing niche areas. In the next section we discuss a different interpretation of engineering research evidenced in engineering journals during the 1950s.

ENGINEERING JOURNALS IN THE 1950s

We turn attention now to the content of engineering journals published in the 1950s because a significant number of papers emphasizing the whole of product system level of development were published during that decade. This section of the chapter is supported by a detailed analysis by the author (Ferris, 2007a, 2007b, 2008). In the context of the present work the point to be drawn from the history of the 1950s publications was the recognition of the achievement of a significant engineering design providing a product or system as being, itself, a research product and worthy of publication in an engineering research venue. This recognition and emphasis receded with the changes in education and engineering research following Sputnik.

Two important trends influencing the understanding of research in engineering occurred during the 1950s. Before the 1950s engineering was taught in institutions with names such as "Institute of Technology", and was only just entering into institutions known as universities. During the 1950s the professional standing of engineering was growing based on the obvious benefits of the rigorous application of complex scientific methods to the achievement of engineering objectives. This shifted the standing of engineering from being concerned with the instantiation of things in workshops and factories, with all the associated connotations of dirty hands, to a field in which there was obvious benefit to be obtained through the application of advanced scientific theory, with the connotations of dealing with generalizable knowledge. This difference was important in the transition of engineering to become recognised as having a proper place in the university. The new view of engineering more closely aligned with concepts associated with the Platonic ideal associated with most of the other professions based on university education. The transition was supported by the obvious advances in engineering which had occurred during World War II as a result of fundamental scientific research into matters enabling the development of technologies which had made a significant difference in the war effort.

Universities were seen as the highest place of learning in most fields. So, during the 1950s when the physical sciences had already been accepted as legitimate fields of university activity, fields such as engineering were trying to be recognised as acceptable. The claim for engineering to enter the university was based on the demonstrable importance of application oriented scientific research in the advancement of engineering achievement of significant new technologies. Application oriented scientific research is research focused on the development of knowledge of the scientific principles associated with particular phenomena which are expected to be useful in applications in the foreseeable future. The pathway to be accepted into the university as a suitable field relied on a field demonstrating credentials as suitably "scientific". The claim to be scientific was supported by following something that appeared to be "the scientific method" in research and teaching in the manner used in the natural sciences. The focus of the natural sciences is to explain how and why things are as they are (March & Smith, 1995) which contrasts with the fundamental concern of engineering with the properties of artificial things (Simon, 1996). The matter of the scientific method will be discussed in a later section of this chapter.

Engineering journals report research related to engineering. Since the 1960s the majority of articles in the majority of engineering journals have reported investigations of particular situations or phenomena which appear to be potentially interesting in the possible engineering of solutions to needs. In this sense the work belongs in engineering journals, which properly contain research which is of interest to those who intend to make artificial things to have a planned effect.

However, the approach to investigation, as reported in most engineering journal articles, is largely abstraction from the context of a solution to a complete need to provide characterisations of particular phenomena, materials or components which may be useful in future engineering work. The post 1960s engineering research paper style is to present the research issue in an abstracted manner, as an issue separated from the practical context which motivated, or any practical application of, the work. In some fields the motivation of projects is to refine the theory which is already in the literature but without the refinement being called for by those who practice in the field.

The research methods which are used include empirical methods in which tests of real instances of the subject matter are performed to build conclusions and simulation and modelling studies in which the properties of a model of a scenario are investigated. Modelling and simulation research can be very valuable because it enables the exploration of the characteristics of a wide range of cases cheaply and conveniently. The value depends on whether the subject matter is amenable to physical investigation or can only be investigated through simulation. In either case the knowledge developed is of a "know that" kind that concerns the properties of the abstraction investigated rather than the behaviour of the subject matter in the context of a real instance.

Another notable change in engineering research papers commencing in the 1960s has been a shift in the balance of authorship of research papers to the university sector from the engineering practitioner sector. This change is associated with the change of subject matter because practising engineer's concerns are with finding means to address needs rather than describing abstractions of things which might be useful in some future engineering project.

In the 1950s the engineering journal literature was rather more diverse. There was greater diversity of authorship, including a significant proportion authored by practicing engineers, and diversity in the classes of subject matter. There were a reasonable number of papers discussing advances in engineering science but there were a large proportion of papers which discussed the development of solutions to particular needs. Some of these papers described solutions to needs addressed by engineers in their engineering work, that is the papers presented theoretical frameworks for particular kinds of analysis that addressed particular practical design needs. For example, the early volumes of *IRE Transactions on Reliability* outlined the theory of reliability as required by engineers who needed to do reliability analysis as part of their design activity. There were also many papers which described the implementation of product or system designs. These papers were written in a manner that would help other engineers confronting related problems to recognise the important issues and approaches to solutions which had been demonstrated successful by somebody else.

Of the two classes paper identified in the 1950s the former class, the theoretical and empirical contributions related to understanding of engineering relevant phenomena are fairly closely aligned to the papers published more recently. However, even here there is a significant difference. The papers published in the 1950s generally provided a clear, brief, explanation of what need in the physical domain had prompted the work which was reported in the papers. The work was not reported in an abstract manner, largely separated from its contribution as part of the solution of a broader engineering problem.

The papers describing the design of complete products and systems are a genre which has become rare in the engineering research literature. The papers describing the design of complete products and systems describe the results of projects in which the act of design created new knowledge that was perceived as worthy of publication as a description of new knowledge. The concept of engineering design as research, creating new knowledge of significant value to the engineering community, in the act of successfully designing a novel solution to a need was accepted in the 1950s, and has since faded from the engineering community.

The difficulty with accepting design as research arose during the 1950s as the engineering community sought to gain university status and consequently shifted its concept of research to that which would produce knowledge of the "know that" kind. The change in the kind of knowledge sought in engineering research resulted from the Platonic ideal, resulting in knowledge of the abstraction of something, rather than the concrete instantiated form of the subject matter, was seen as necessary to gaining a place within universities. The result has been an emphasis on discovery of generalisable knowledge of the "know that" kind in engineering research. The effect has been to separate the academic and practice communities so that, whilst an engineering degree is a prerequisite for entry into either the academic or engineering practice communities the two communities and career paths then diverge.

1950's EXAMPLES OF DESIGN AS RESEARCH

We now discuss three examples of papers demonstrating design as research dating from the 1950s. These are only samples from amongst many available from the decade. The first of our studies concerns the development of a photographic data logger which included both time and magnitude grids (Riblet, 1956). This paper commences by reviewing existing data logging and plotting instruments, noting that there are two classes of recorders. The two classes are those that directly write on a paper chart, useful for low-frequency signals, and those producing a photographic record through reflecting a light beam from a mirror, capable of responding high-frequency signals. The latter class do not include calibrated taxis and grids for the timescale and magnitude resulting in a very slow process of reducing data to a usable form.

The authors introduced their project objectives. They set out to develop equipment which could simultaneously plot a function on the time grid and make the chart produced in the data plotter the final data form. Their product simultaneously plotted the function being measured; the time grid; and a magnitude grid. They recognised that achieving these objectives would require design of three significant elements. After describing the goal the paper describes the design solutions for each of these major elements. The description was a verbal description of the elements as built and the rationale applied in the design. The paper concludes with an evaluation of the instrument developed in terms of the quality of the plots produced and an evaluation of the labour saved.

The second study concerns the development of a plasmatron tube (Johnson & Webster, 1952). Existing vacuum and gas tubes provided either high impedance, continuously controllable, low current or low impedance, uncontrolled, high current operation. The absence of a tube type capable of providing low impedance, high and controllable current operation presented a difficulty for designers in certain applications.

The tube designed was described using words to explain the principal by which worked, photographs to show its physical construction and test results. In addition the authors presented formulae which would assist in design of the components to achieve desired performance and a theoretical description of how the tube performed. The paper provided the information necessary for another engineer to design a similar product and the principal performance characteristics graphs to enable readers to understand its performance characteristics. This paper showed the design of a novel component type with the information necessary for users to use it or for others to design similar components.

The third case concerns the design of a very high power pulsed magnetron tube (Okress, Gleason, White, & Hayter, 1957). These authors provide detail of the electrical characteristics of the tube which they were required to build, noting that the specified parameters were very ambitious. The authors identified a number of major technical challenges. The technical challenges included achieving the required performance and achieving reliability that would make the component useful. The work commenced in 1946 and design was completed in 1953 with the paper being published in 1957, including an account of the reliability achieved.

The methods of addressing the technical problems resulted in development of appropriate materials. The paper describes the scientific experiments performed to demonstrate the suitability for purpose of the materials found. A further problem identified and solved was the difficulty of manufacture of such a large tube.

It is clear that the product was very challenging. The purpose was not explained but much else about the design was described. The performance and test results for the completed tube were provided showing successful completion of the project. What is significant is that this paper shows how the authors successfully satisfied their objective and solved a number of significantly challenging problems.

Each of the three papers discussed in this section demonstrated the successful design of a product which was inherently very challenging showing how such a design project can satisfy a novel need in a suitable way and how design activity may demand discovery of various kinds of solution to many distinct needs. The solution of the various needs that together comprise the project may involve a variety of research methods and lead to discoveries of various kinds.

DESIGN AS RESEARCH METHOD

The idea of the scientific method of how scientists proceed with research is a myth concerning how observations are transformed into knowledge. The common descriptions of the scientific method start with the proposal of an hypothesis which will be tested through the collection of appropriate empirical data. By nature of the case the empirical observations can only refute the hypothesis but can never prove it. This situation leads to several problems:

1. Often the research question is distorted so that the researcher can create an hypothesis which can be tested given the resources available.
2. Often the researcher's belief about the situation is reconstructed as a null hypothesis so that when the null hypothesis is refuted the researcher feels justified in asserting that their belief about the subject matter is supported.
3. The method, as described, does not provide an explanation of how the researcher can develop knowledge in a field which does not already have a body of work which can be used to construct formal hypotheses.

4. The result of such research is conclusions about particular hypotheses which represent abstractions from the context in which the observed phenomena exists or any situation in which the knowledge might be useful. The objective of such research is the development of knowledge for its intrinsic value, not its instrumental value.

The result of research performed using an hypothesis testing approach is the creation of scientific knowledge describing certain aspects of the matters which are being investigated. This knowledge is often obtained in a situation where the researcher's objective is to obtain information about a situation which appears interesting for the intrinsic value of the knowledge sought. The impact of orientation to the intrinsic value of the knowledge is that the questions asked and the termination point of the research project are determined based on the expected achievement of the knowledge which is perceived to have an intrinsic value. The consequence is that the knowledge obtained may not be relevant to any application. The problem of distortion of the research questions in order to have questions which are amenable to answer by the standard method frequently results in good work which is difficult to apply.

The nature of the research process discussed above is that it provides knowledge about questions which are abstracted from their context. Abstraction is often necessary in order to construct the problem in a form which can be addressed using this approach. However, abstraction by its nature removes the matter of investigation from its context resulting in knowledge which is largely context independent. The difficulty of abstracted context independent knowledge is that it is difficult to integrate such knowledge into an arrangement which applies that knowledge to achieve a desired outcome. The difficulty in applying scientifically derived knowledge to achieve a desired outcome is a consequence of the different kinds of knowledge, as discussed above.

Scientifically derived knowledge concerns the development of a generally valid theory which explains the observations. As such it concerns knowledge of the "know that" kind. That is, scientifically derived knowledge is concerned with description of things, transforming our awareness of those things from perception of the state of nature to description of a theory that explains the state of nature. The description of the theory is, inherently, and explicit articulated form. When someone desires to use knowledge and to integrate elements of knowledge to create some meaningful product the knowledge required is not just of the "know that" form describing the elements which must be integrated. The act of integrating the fragments of knowledge to achieve a useful outcome is an act of constructing knowledge in either the "know how" or "knowing" forms. The dynamism of the knowledge created in the act of application to the solution of a need occurs because in the act of integration of the elements of knowledge that described the abstracted phenomenon of the elements are transformed from knowledge about abstraction is to means to achieve an end.

We consider now the fundamental difference between science and engineering. Science seeks knowledge which enables the description of a state of affairs which already exists. Engineering is concerned with action which leads to desired outcomes. Engineering has two foci of action: the design of the product system itself; and the engineering organization and process which achieves the production focus. In relation to the engineering process some perform research of a prescriptive kind, as described by a number of design scientists, whose interest is to determine the processes that would be prescribed in order to achieve a desirable design (March & Smith, 1995; Walls, et al., 1992). In the development of a prescriptive theory, that is a theory which is designed to instruct future action, the researcher must be careful to consider the impact of the context in the situations observed, which are used to develop the theory, so that that context does not produce inappropriate prescription for other contexts (Dorst, 2008).

In relate to the former objective of engineering, the development of suitable solutions to needs, leads to work continuing until the need is satisficed (Simon, 1996). If the need is routine we may be able to rely on precedence to guide a new design. If the need is very ambitious we might need to perform fundamental research to discover methods to solve the problem. In the case of routine design there is little that is particularly new that is discovered in the design process, since the design space is constrained to lead to an incremental variation of existing methods to achieve a suitable solution. In the case of design of novel products and systems there will be many discoveries made as the engineer seeks to integrate existing related knowledge of the various elements into a solution to the need. Another difference between science and engineering is that engineering is concerned with achieving a sufficient solution, which satisfices the need, rather than the completely satisfying theory sought by the scientist.

The engineer's goal is to make something which is useful. Usefulness is the characteristic that determines if an engineered product is successful and, by extension, the success of the engineer (Hopkins, 1954). To be useful a product must suit its purpose and intended situation of deployment, its *sitz im leben*.

To satisfy the opposite for the intended *sitz im leben* evaluation criterion the engineer may need to do significant research to develop understanding of the need. This requires significant care to ensure that the questions asked in the investigation do not lead to a particular set of findings which lead ultimately to the circular reasoning that the investigator finds what they presumed to be present by the nature of the question which they asked (Dobson, 2001) The process of designing a novel product, or a product using a different technology than has previously been used, to address that need will produce significant new knowledge of how that need can be addressed. Therefore, design is a means of creating new knowledge concerning the problem space and means of addressing the problem, and so a methodology for research.

The differences between engineering and science imply different epistemologies and possibly different ontologies. Both believe that there is a real world with which practitioners interact. Neither the engineer nor the scientist is a post-modernist, but rather both are significantly influenced by Positivist reasoning. As such both believe we can observe the world and make real observations and that multiple people could make the same observation. Both are Critical Realists who believe that observations are affected by the circumstances in which they are done but both believe that the object which prompted an observation is real and that differences in what is seen are the result of perturbations caused by the observation situation and differences of view point experienced by different observers. Recognition of this problem has led physical scientists and engineers to develop methods to compensate for variations of observation caused by the point of view. Both the scientist and engineer are positivists because they believe that there is real stuff.

The difference between the positions taken by the scientist and engineer is that a scientist is concerned with creating universally true knowledge and, in contrast, the engineer's concern is the delivery of apposite satisficing solutions to needs, which results in the engineer having a distinctly specific view of knowledge and discovery. To this end it is necessary for the engineer to have case specific knowledge. The engineer needs both general scientific knowledge and case specific knowledge. The engineer will also use phenomenological knowledge, that is knowledge which describes something repeatable may be useful even in the absence of a substantial theory to explain the observations. A lot of successful engineering work has been based on phenomenological knowledge, such as the engineering work based on graphical representations of relationships of variables where the relationship was too complex to express mathematically or numerically at the time the engineering work was done.

KNOWLEDGE DEVELOPED THROUGH DESIGN

Design and construction of a solution to a need demonstrates that it is possible to achieve a solution to address the need. That knowledge, alone, is not very exciting and would not be a satisfactory finding of a research project because it provides no foundation for further action. What is exciting as research is the development of understanding of how the need was addressed so that in future similar, but different, cases we have knowledge that will guide our approach. In this, useful research through the activity of design develops a theory which links design requirements with the solution developed and the method of developing the design (Markus, Majchrzak, & Gasser, 2002). Contrary to Hickey (1960), who asserted that "engineering answers the hypothesis that it is possible to build something", the purpose of engineering design is not to answer the hypothesis that it is possible to build a solution to the particular need, it is about coming up with an apposite satisficing solution to that need.

The engineer's goal is to produce a suitable solution to the need. Design based research determines the characteristics of a suitable solution, and carries through at least one design to a state in which it satisfices the need. Design based research in engineering concerns understanding what constitutes a suitable solution and determining whatever is needed to achieve such a solution. Hickey's view is a force-fit of an engineering activity into a scientific conceptualisation of research, which is inappropriate. It is notable that Hickey's view was published in 1960, when engineers were seeking to present and promote engineering as "scientific".

A point of difference between the design based research methodology and the scientific methodology is that design produces an instance of a solution to the need. That is, design produces a specific instantiation related to a specific problem. It is a reasonable to expect that there might have been other ways to address the need. And, we cannot prove that the solution offered is the best solution but rather, only an effective or satisficing solution. In contrast a scientist seeks to demonstrate that knowledge created is true, and so aims to make absolute assertions. Consequently, scientists are dissatisfied with the need to specify caveats on the quality or generalizability of the knowledge that they obtain.

Whilst the scientist cannot prove that the knowledge created is true their intention is to create universally applicable knowledge. In contrast the engineer's solution is never claimed to be universally desirable nor the best possible solution, but rather the best that the engineer was able to produce in the circumstances. Design yields background knowledge about components, materials and their behaviour which impact the product, but may also be useful in other cases.

CONCLUSION

Research is generally recognised as an activity which creates new knowledge. The discussion in this chapter has concerned the manner in which the action of designing novel engineering solutions to needs is a research process, contrary to much popular opinion. This chapter has examined the underlying conceptualisation of research embodied in the form of research which has become common in engineering since the 1960s, which is a form of research similar to that performed in the various science disciplines, and largely based on Positivist principles.

Underlying the conceptualisation of research is the question of what is seen as constituting knowledge. Research methods associated with the development of objective and generalizable knowledge are designed to produce scientific type understanding of extant things. This contrasts with the fundamental nature of engineering which concerns the development of solutions to needs. Therefore engineering has three levels of specificity. The need is specific and must be appropriately understood. The architectural approach to the design of a solution presents decisions about specific methods of instantiation of a solution to the need. The third layer of specificity in engineering design is the particular design which is used to realise the architectural approach which has been decided for the solution. Each of these three specific matters results in discovery related to specific ways of achieving solutions to the needs. The effectiveness of solutions which are offered may be evaluated but it is not possible to determine if any solution which is offered is the best possible solution or the best solution that could have been offered at a time or the best solution that could have been offered for a certain resource expenditure or according to some other measure of desirability. All that can be determined is that the solution satisfices the need when evaluated across the range of criteria determined as appropriate evaluation criteria. Therefore the kind of knowledge created by design as a research methodology is fundamentally different than the kind of knowledge created by science oriented research methodologies.

The knowledge created by science oriented research methodologies is desired for the intrinsic value it is believed to have. The knowledge is expected to be objective and generalizable and will be of some explicitly expressible form, consistent with the "know that" conceptualisation of knowledge. The knowledge created by the design methodology is desired for its instrumental value; the ability to achieve a solution to a problem that it enables. Knowledge generated through design is specific to the need addressed and the design approach used. The knowledge generated through design is of the "know how" form and may contribute to knowledge of the "knowing" form as experience with a number of design projects enables an engineer to develop better judgement about appropriate design strategies. The knowledge developed through design includes knowledge related to the achievement of the product which is designed and may contribute to knowledge of the design process as means for the performance of future design projects.

REFERENCES

Biggs, J. (1999). *Teaching for quality learning at university: What the student does*. Buckingham, UK: Society for Research into Higher Education, Open University Press.

Bucciarelli, L. L. (2003). *Engineering philosophy*. Delft, The Netherlands: DUP Satellite.

Department of Education Training and Youth Affairs. (2001). *Factors associated with completion of research higher degrees*. Department of Education Training and Youth Affairs.

Dobson, P. J. (2001). The philosophy of critical realism - An opportunity for information systems research. *Information Systems Frontiers*, *3*(2), 199–210. doi:10.1023/A:1011495424958

Dorst, K. (2008). Design research: A revolution-waiting-to-happen. *Design Research*, *29*(1), 4–11.

Ferris, T. L. J. (2007a). *Some early history of systems engineering - 1950's in IRE publications (part 1): The problem*. 17th International Symposium Systems Engineering: key to intelligent enterprises. San Diego, CA: International Council on Systems Engineering.

Ferris, T. L. J. (2007b). *Some early history of systems engineering - 1950's in IRE publications (part 2): The solution.* 17th International Symposium Systems Engineering: Key to intelligent enterprises. San Diego, CA: International Council on Systems Engineering.

Ferris, T. L. J. (2008). *Early history of systems engineering (Part 3) – 1950's in various engineering sources*. INCOSE International Symposium 2008. Utrecht, The Netherlands: International Council on Systems Engineering.

Ferris, T. L. J. (2009). *On the methods of research for systems engineering*. 7th Annual Conference on Systems Engineering Research. Loughborough, UK.

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, *8*(5), 312–335.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in Information Systems research. *Management Information Systems Quarterly*, *28*(1), 75–105.

Hickey, A. E. Jr. (1960). The systems approach: Can engineers use the scientific method? *IRE Transactions on Engineering Management*, *7*(2), 72–80. doi:10.1109/IRET-EM.1960.5007541

Hopkins, M. (1954). Human relations in engineering management. *Transactions of the IRE Professional Group on Engineering Management*, *2*, 16–27. doi:10.1109/TPGEM.1954.5010170

Johnson, E. O., & Webster, W. M. (1952). The plasmatron: A continuously controllable gas-discharge developmental tube. *Proceedings of the IRE*, *40*(6), 645–659. doi:10.1109/JRPROC.1952.274057

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in Information Systems research. *Information Systems Research*, *14*(3), 221–242. doi:10.1287/isre.14.3.221.16560

March, S. T., & Smith, G. F. (1995). Design and natural science research on Information Technology. *Decision Support Systems*, *15*, 251–266. doi:10.1016/0167-9236(94)00041-2

Markus, M.-L., Majchrzak, A., & Gasser, L. (2002). A design theory for systems that support emergent knowledge processes. *Management Information Systems Quarterly*, *26*(3), 179–212.

Nissen, M. E. (2006). *Harnessing knowledge dynamics: Principled organizational knowing & learning*. Hershey, PA: IGI Global.

Okress, E. C., Gleason, C. H., White, R. A., & Hayter, W. R. (1957). Design and performance of a high power pulsed magnetron. *IRE Transactions on Electron Devices*, *4*(2), 161–171. doi:10.1109/T-ED.1957.14222

Riblet, H. B. (1956). A simplified automatic data plotter. *IRE Transactions on Instrumentation*, *5*, 34–43. doi:10.1109/IRE-I.1956.5006999

Ryle, G. (1948). Knowing how and knowing that. *Proceedings of the Aristotelian Society,* (p. 46).

Simon, H. A. (1996). *The sciences of the artificial*. Cambridge, MA: The MIT Press.

Walls, J. G., Widmeyer, G. R., & El-Sawy, O. A. (1992). Building an Information System design theory for vigilent EIS. *Information Systems Research*, *3*(1), 36–59. doi:10.1287/isre.3.1.36

KEY TERMS AND DEFINITIONS

Design: A human activity which combines existing knowledge and things to generate an output which is desired.

Engineering: A human activity centred on providing appropriate solutions to needs.

Know How: A form of knowledge which concerns being able to do particular itemised actions.

Know That: A form of knowledge which can be explicitly stated, and so can be learned and taught in the form of those statements.

Knowing: A form of knowledge related to the fluent and appropriate performance of action.

Knowledge: That which one has a suitable foundation for being assured is a true or appropriate description or method of interacting with the world or a part thereof.

Research: A human activity which creates new knowledge.

Science: A human activity centred on seeking to understand observed phenomena, things and situations with the purpose of providing accuratedescriptions and explanatory theories ofthe things observed.

***Sitz Im Leben*:** Situation in life, following usage in Theology which originated with Hermann Günkel.

ENDNOTE

[1] This was stated in the Master of Engineering by Research and Ph.D. regulations of the University of Adelaide in the late 1970s.

This work was previously published in Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems, edited by Manuel Mora, Ovsei Gelman, Annette L. Steenkamp and Mahesh Raisinghani, pp. 389-402, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 95
Engineer-to-Order:
A Maturity Concurrent Engineering Best Practice in Improving Supply Chains

Richard Addo-Tenkorang
University of Vaasa, Finland

Ephrem Eyob
Virginia State University, USA

ABSTRACT

General Engineer-to-Order (ETO) design of product capacity projects among many others, includes design for large electric machine, huge centrifugal pumps, diesel/natural fuel power plant engines, steam turbine, boiler, ship power, et cetera. ETO is basically a product development process, which starts with a product specification and finishes with an engineering design as its deliverable. It rarely includes manufacturing processes. The main drawback is with issues concerning its long lead-time. Research shows that an excessive lead-time is more often than not caused directly or indirectly by factors related to the design phase. This chapter thus, endeavours to introduce a best practice concurrent approach for reducing the lead-time at an engineer-to-order product design/development stage by seeking to integrate business information technology systems in the design and operational phases. It also introduces a new concurrent best practice approach by way of seeking to integrate other related business systems, e.g., (Enterprise Resource Planning (ERP)) such as (Enterprise Service Architecture (ESA) application processes with Service-Oriented Architecture (SOA)) as a platform for applications and processes for effective communication. Furthermore, the chapter presents and discusses a model of classical concurrent engineering (CE) ETO operational process. ETO key elements, ETO success factors, and series of state of the art ETO classical ERP engineering design tools, as well as the "best practice" product life cycle are all discussed.

DOI: 10.4018/978-1-4666-1945-6.ch095

INTRODUCTION TO ETO FROM A CONCURRENT ENGINEERING PERSPECTIVE

Industrial manufacturing companies producing engineered products on an engineer to order (ETO) in a concurrent engineering approach (i.e. partners from various divisions in the organizations) have traditionally found business opportunities. Through the design and product development expertise and an ability to respond to demands for customisation with improved product performance, customers are increasingly seeking for more affordable prices and a much reduced lead-times, which also require improved manufacturing efficiency. These companies are thus being driven to improve the integration of the design, manufacturing and procurement functions (Cameron, and Braiden, 2004). In order to facilitate this improvement in the integration of the various partners involved, and to improve the overall product time to enhance and better define the product as well as properly document the design process, a concurrent engineering process needs to be adopted in the supply chain management (SCM). Concurrent engineering process which very much operates in the same vain as a collaborative engineering process, is a process in which the appropriate partner organizations are committed to integrating their system processes to work interactively. Thus, to visualize, develop and implement engineered products and programs that correspond with their pre-determined objectives. Concurrent engineering in this sense is moderate collaborative engineering approach that is much more focused on the engineering product design process of concurrent cross-functional partner organizations together in order to create products that are much better, much affordable and with a much reduced lead-time.

Engineer to order products may require a specific set of item numbers, bills of material (BOM), and routings; these are usually complex with long lead times. Customers are also normally heavily involved throughout the entire design process and manufacturing process for an engineer to order a product. It is usually limited to an engineering design process which involves the tasks of engineering analysis, concept design, architectural design, detailed design and manufacturing process design. It does not include the manufacturing phase of materials acquisition, fabrication and assembly. In a traditional ETO operation, product functionality is the major design focus (Chen, 2006).

The product design process phase takes enormous time of the entire process lead-time. Thus, the new and improved ETO – CE process flow model approach proposed in this chapter seeks to reduce the long lead-time by its CE approach of all the partners as well as communicating and exchanging product design information on a common platform. This new improved approach as compared with the classic ETO process flow approach will bring all partners in the design supply chains together both locally and globally by using a system oriented architecture (SOA) – extensible mark-up language (XML) communications on a common platform across the world-wide-web. Thus, providing an opportunity for even different design teams in different organizations, with different SCM information systems (IS) to link-in easily and faster to resolve any design issues. Thereby, reducing any delay in design processes, which is the cause of long lead-times in classical ETO processes. ETO – CE approach in improving product design supply chains enhances an industrial manufacturing organization's ability to gain effective and efficient control of the upstream through to the downstream of their ETO product design processes and assume a dynamic concurrent approach to supply chain rather than a reflexive to lead-time threats. The follow section gives more elaboration of how ETO – CE integration into supply chains could improve the chain.

Supply Chain - ETO - CE Process Integration

ETO's successful integration into manufacturing supply chain is by efficiently and effectively incorporating the various engineer-to-order activities and processes from the upstream in the supply chain (suppliers) through to the downstream (end customers). Considering the current global competitive pressures, manufacturing industries have followed through with series of business strategies and different approaches in their bid to enhance the value for the end customer (Kim, 2005; Chopra and Meindl, 2001). Industrial manufacturers who succeed globally continue to adapt, develop and improve their supply chain network value as the market condition becomes more progressive (Kumar et al., 2006; Reichhart, and Holweg, 2007).

The significance of speed in product development activities and processes, lean manufacturing practices, quality, cost and flexibility in the supply chain should be well supported and managed by advanced infrastructures in supplier sourcing, logistic management, information systems and technologies as well as very good customer relations management (Gunasekaran et al., 2008). Thus, value in ETO products is realised in product development of the supply chain, mostly in a concurrent approach through collaborative partnerships and integrations that recognize contributions from the product development process, procedure, information, finance, management of knowledge, innovation and relationship management (Barber, 2009). Therefore, the approach of successfully integrating ETO methodologies enhances supply chain flexibility and value. This reflects the current state of embedding process improvement into the supply chain operations of product development a as being proactive in managing supply-demand fulfillment concurrently (Lee, 2004; Ketchen et al., 2008; O'Marah and Hofman, 2010; Porter, 1990).

Integration strategy programs, applications and processes involving all stakeholders in product development (i.e. customers, suppliers, marketers, accountants, design engineers, production engineers, manufacturing engineers, etc.) during the capacity management, enhances the ETO supply chain to a shorter delivery lead-time on customer specification and orders (Swafford et al., 2006). Customer relationship effect in ETO methodology integration in the supply chain increases understanding in customer demand, managing complaints and improving customer satisfaction which critical responsive element through the ETO value chain (Li et al., 2006; Holweg, 2005). According to Zhou and Benton (2007) quality information sharing on a common and assessable platform among stakeholders enhances ETO supply chain practices such as just-in-time (JIT) and lean lead-time delivery. Thus, the capability to collect, disseminate various data, information from customers across ETO supply chain nodes concurrently will enable internal manufacturing strategies to respond effectively to the needs of customers. This paper presents a maturity concurrent engineering "best practice" approach in ETO methodology in product development. An improved CE- ETO model proposed in the subsequent sections in a concurrent approach is expected to reduce the lead-time making ETO methodology leaner by also successfully integrating enterprise resource processes and applications in the ETO supply chain.

Thus, new trends reunite technical and non-technical disciplines such as product engineering partners, marketing/brokers and accounting/financial controllers. This collaboration aims at satisfying the customer. Therefore, the product design partners in industrial organizations must also collaborate in defining and analysing the product specification aimed at satisfying the end customer. Hence, the approach to ETO process, concurrently in a collaborative approach, mainly dwells with the product design aspect. ETO product and the engineering process lead-time is well managed and compensated for in the product design stage which turns to take a chunk of the process time. Engineer to order is a manufacturing process whereby finished products are built to unique customer specifications. These engineering product design processes are elaborated in detail in the following sections from engineering analysis to manufacturing process design.

Engineering Analysis

Engineering analysis is the aspect of engineering product design, which involves technical investigative processes and theories to reveal the properties and circumstances of an engineering system, product or concept design/structure. Engineering analysis also involves the tool and methods required in the technical investigative processes or theories. This when successfully investigated could lead to an engineering design or ironically to a simply failed process that could always be re-analysed to find out where the failures lay in the process? Therefore, this analysis must always be carried through before proceeding to produce a concept design of an ETO product to rectify the huge cost which comes with failures in the engineering systems and process design.

Concept Design

Concept design is the very first stage of an engineering design work where all the engineering drawings are mainly centered. These concept designs consist of very basic engineering 2D drawings and sectional views. The 2D engineering draws and sectional views are later developed along the process line to a more detailed 3D engineering draws with their sectional views as well. Concept design in an engineering design process has certain specific stages of conceptual design work stages that are required to transfer information into the needed requirements for the detailed engineering design of a specific engineering product. These stages include a description of the general concept, definition of the requirements or specifications of the concept plan and description of what the concept is aimed at realizing at the end as well as lists of activities for the concept. A good conceptual design comprises of the creation of a proposal or an initiative, the investigation of that proposal, or an initiative into a representation of the proposal. On this note, conceptual design can range from basic blueprint line diagrams to structural 2D engineering drawings. Thus, if the concepts are not clearly drawn, the proposal will not be completely understood, and the result potentially will be compromised.

Architectural Design

Architectural design means often the same as engineering design. It is the application of engineering design standards to engineering product design works and bill of materials (BOM) of the detailed design. Architectural design is the state in engineering product design where the conceptual design is further developed, modified or processed in a more detailed design capable of being sent to the next stage of product design.

Detailed Design

This stage in the engineering product design stage is where the detail engineering design is achieved by the design team. By using both the requirement specification (conceptual design), and the architectural specification (architectural design), provided in the previous stages of the design process. The outcome at this stage of the process will be a detailed design which will be featured in detail, describing the interfaces and functions provided by the engineering design product. This detailed design will then serve as the basis for the manufacturing process stage.

Manufacturing Process Design

This is the stage of the engineering product design where the detail design is processed for manufacturing. Material request and acquisitions are determined at this stage for fabrication and product assembly. Thus, the process and technology analysis, assembly process planning and component process planning in the detail design stages are all implemented at this stage.

The following section introduces some of the benefits for CE approach in ETO to improve the design chain and reduce the lead-times. It discusses existing research on these industrial manufacturing approaches and also their collaboration and integration successes as well as the various stages of a product life cycle (PLC).

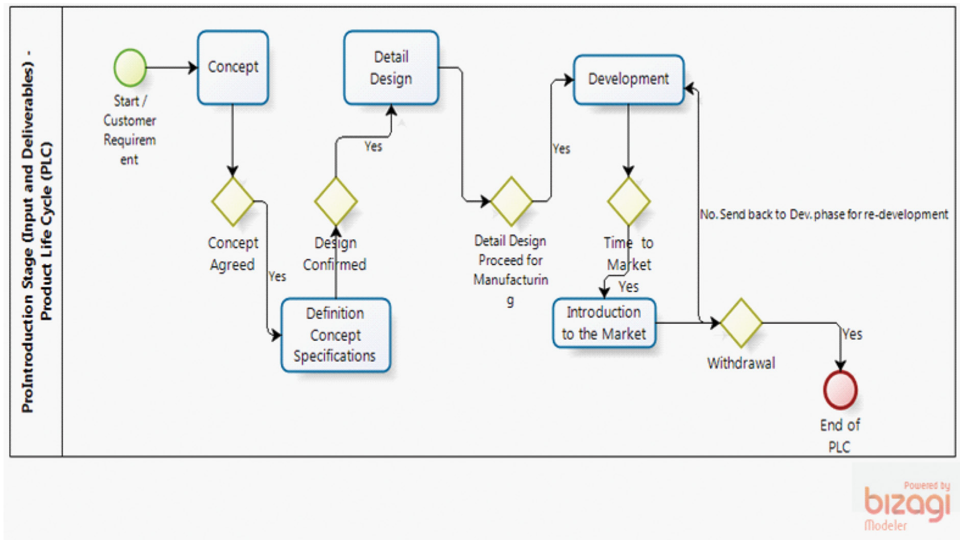
BACKGROUND REVIEW

Over the past decade industries in almost all markets have been facing a boosting level of competitiveness. There are many reasons for this, but most of them can be followed to some of the following principal trends: shortening product life cycles, globalization of the market, rapid technological changes, environmental issues, and higher complexity of products, customers demanding products with more features, higher quality, lower cost, and demand for more and more customized products. Concurrent Engineering (CE) is an integrated product development approach that emphasises the response to customer expectations by producing better, cost effective and much faster product development process. It also supports multidisciplinary team values of co-operation and trust; thus, sharing and exchanging required knowledge and information in a manner that will enhance decision making processes and also emphasis on simultaneous consideration during the design stage and all the other Product Life Cycle (PLC) aspects of the product development. PLC is the progress of a product from its concept through to its development and introduction to the market and then its withdrawal from the market or final disposal. One of the most salient means to reduce development time is through the use of "concurrent engineering." Concurrent engineering is defined by the Institute for Defence Analysis (IDA) as "the systematic approach to the integrated concurrent design of products and related processes, including manufacture and support".

Thus, PLC management confronts the need to balance fast response to changing consumer demands with competitive pressure to seek cost reductions in sourcing, manufacturing and distribution. It needs to be based on a close alignment between customer-facing functions (e.g. marketing, sales, customer service) and supply functions (e.g., purchasing, manufacturing, and logistics) (Combs, 2004; Conner, 2004; O'Marah, 2003). Hence, Product life cycle (PLC) management has become a strategic priority in many company's boardrooms (Teresko, 2004). According to Yassine, and Braha, (2003); Concurrent engineering (CE) is an engineering management philosophy and a set of operating principles that guide a product development process through to an accelerated successful completion. In general, CE values rely on a single, but powerful, principle that encourages the incorporation of the later stages of production concerns into the upstream phases of a development process chain such as during the engineering-to-order (ETO) phase of a product development, etc. This would lead to shorter development times, thus, improving product quality, and lowering development–production costs.

Concurrent engineering is hereby aimed at the timely availability of critical design information to all the design development participants. For most intricate engineering tasks all significant information required by a specific design development team cannot be completely available at the start of that task. Therefore, CE requires the most of such information and the ability to share and communicate useful information on a timely basis with the right design experts. The concept of concurrent engineering (CE) has been known for quite a while now, and it has been widely recognized as a major enabler of fast and efficient product development. This chapter examines the extent to which CE "best practices", as obtained

Figure 1. Introduction stage (input and deliverables) - Product life cycle (PLC) - Input and deliverables (BizAgi process modeller)



from a broad literature review, are being used effectively in companies globally. Some of the positive impact of formal CE programs is proven by previous research (Portioli-Staudacher, *et al.*, 2003). Figure 1 illustrates the stages of a product life cycle (PLC) highlighting the inputs and deliverables in one completed flow process with intermittent decision loop at each milestone stage of the flow process.

As the product development stage is wiry in market size but there is a growth supply chain, it is imperative that substantial research and development costs be incurred in getting the product to this stage. In addition, marketing costs may be too high to just test the market or undergo commencement promotion and set up distribution outlets. Therefore, it is highly unlikely that industries will make profits on products at the development stage when rigorous work has not been duly than to fulfil all the required customer requirements. Products at this stage have to be carefully monitored to ensure that they start to grow and pick in the market. Otherwise, the best option may be to withdraw or end the product. The need for immediate profit is not a pressure as the lack of it is an expectation at this stage. The product is promoted to create awareness of the market as well as satisfying customer requirements. Thus, if the product is not competitive, an initial pricing strategy is employed to build pricing momentum to maximise profits. Limited numbers of product will be available in few outlets of a distribution chain. Therefore, development stage should encompass a number of activities that will include:

- **Concept:** Overview of the customer requirement that an opportunity seeks to address, supported by evidence of market need.
- **Definition:** High-level definition of customer requirements and analysis of a business opportunity.
- **Design:** Analysis of customer requirements creating project plan and detailed product specification.
- **Development:** Data and software (CAD/CAM) development.
- **Development Testing:** Testing of the product against pre-defined test schedules

to ensure satisfactory performance against customer requirements.

- Development of pricing.
- Development of user guide
- Introduction of the product to the market (Time to Market). etc.

Product lifecycle management and the efficiency of information reuse relies on the definition and management of corresponding information and communication between various product design teams in various manufacturing organizations. Collaborating and sharing information concurrently in the ETO supply chains ensures the consistency and traceability of product information throughout the product lifecycle. The sales-delivery process of engineer-to-order (ETO) products presents a great potential for design reuse, i.e. the reuse of previously validated design solutions in the design of new product variants according to customer-specific requirements (Brie`re-Coˆte´, *et al,* 2010). Much as lean manufacturing (LM) in manufacturing organizations is currently enjoying its second prime, manufacturers in several industries are implementing lean practices to keep pace with the competition and achieve better results by reducing product design lead-times significantly Demeter, and Mateusz (2011).

The principles of concurrent engineering (CE) in the same vein as that of lean production has enabled manufacturing organisations to significantly improve their competitiveness. The application of CE principles has enabled many organisations to simultaneously improve productivity, quality and customer service (Riezebos, *et al,* 2009). Similar benefits have been achieved through the integration of information technology (IT) in ETO processes. Thus, integrating IT and ETO – CE principles are seen to be interdependent and complimentary by some; whilst others have seen that as the approaches as being mutually exclusive. This section thus, presents reviews of the use of some ERP systems could collaborate within the ETO processes to concurrently improve the supply chain. The following section elaborates further on the trends and perspective in ETO new product introduction and development (NPI/D). In order to meet the ever increasing customer demand and preferences is well documented. Therefore, customisation has been stimulated as a source of competitive advantage for ETO industrial manufactures. Despite these factors, most of the researched work in operations and supply chain management has side-lined the need for improvement of the ‘engineer- to-order’ (ETO) supply chain sector (Gosling, and Naim, 2009). Hence the following section will highlight a few of the essence of the global trends and perspectives in ETO new product introduction and development supply chains.

CE Global Trends and Perspectives in ETO

According to Tennant, and Roberts, (2000), an effective New Product Introduction and Development (NPI/D) process, which is a concurrent process, can enhance an organisation’s competitiveness. This can be done by compressing product development lead-times, and enabling the upstream and downstream processes in a supply chain to be considered when taking decisions at the product concept design phase. The application of Concurrent Engineering (CE) (or Integrated Product Development [IPD]) is gradually becoming the norm for developing and introducing new products to the market place (Ainscough and Yazdani, 2000). However, the degree to which companies have implemented it and the amount of success varies (Ainscough and Yazdani, 2000; Balbontin *et al.*, 2000). Many of the companies competing today in international markets consider new product introduction and development (NPI/D) as an important factor for achieving sustainable competitive advantages. Both academics and industrial managers alike are constantly searching for methods and practices that will allow them to improve the organization and management of their NPI/D processes and boost their effectiveness. The average success rate of NPI/D projects today is approximately 60% (Cooper, and Edgett, 2003). The aim of this research is to 1. shorten new product development times, 2. achieve more efficient developments, and 3. superior products.

Manufacturing companies have re-systematized their NPI/D processes and have moved from a sequential path, in which there is a negligible interaction among the departments involved and the activities required to develop the product which is carried out sequentially, towards an integrated path, known as concurrent engineering (CE) in which the activities overlap and all the departments collaborate from the beginning. This new organizational design has helped companies to improve their performance by leading to lower costs, higher quality, major knowledge creation and shorter product development times (Barba, 2001; Umemoto, *et al.,* 2004). All of this has raised competitive skills of companies. Hence, the aim is to avoid continuous setbacks and the other problems that arise with the traditional approach, improving NPI/D performance. This new practice tries to speed up the process, increasing flexibility, adopting a more strategic perspective with more sensitivity to change in the environment, solving problems through teamwork, developing diverse skills, and improving internal communication (Barba, 2001). To achieve the above mentioned objectives, CE is based on three basic elements (Koufteros, *et al.*, 2001):

1. Concurrent work-flow
2. Early involvement of all participants and groups contributing to product development
3. Team work. In other words, CE is the early involvement of a cross-functional team to simultaneously plan product, process and manufacturing activities and mentioned earlier in the previous paragraphs.

Many studies demonstrate that CE can successfully solve the typical problems of traditional. NPD, leading to clear improvements in quality and marked reductions in development time and costs (Calantone and Di Benedetto, 2000; Herder and Weijnen, 2000; Barba, 2001; Koufteros, *et al.*, 2001). On the other hand, further recent research also has also revealed that, the use of CE on its own does not always lead to positive results and that success in improving innovation capabilities depends on the context, in which CE is applied. That is, on the prevailing competitive and technological circumstances (Valle, and Va´zquez-Bustelo, 2009). Therefore, a conclusion is reached that the scale of a vagueness and intricacy present in the process of innovation may moderate the effectiveness of concurrent NPI/D characteristics on performance. Hence, the matter to be considered is not, whether CE is a mechanism for improving performance in the introduction of new products but, rather, under what circumstances such as improvement can be accomplished. It seems, however, that, in spite of many research efforts; a consensus is still lacking and that there are many empirical disagreements. This lack of agreement is the reason for this review of the global trends and perspective of CE.

According to Campbell, and Mohun, (2007), industrial manufacturing company endeavour to create an advanced core-system analytical solution that integration processes across their manufacturing industry. Thus, this will enhance the efforts by industrial processes by reducing lead-time variability and minimizing the transition times to achieve performance consistency. Therefore, to achieve this kind of core-system integrated product design and manufacturing collaboration, service-oriented architecture (SOA) turns out to be the most preferred systems' application platform. SOA is the most suitable systems' application platform for the recommended Enterprise Service Architecture to enhance this integration process effectively and efficiently. Further to this analytical leverage of CE system applications and processes; this collaboration has also been accordingly confirmed by Valle, *et at.*, (2009) research on 'Concurrent engineering performance: Incremental versus radical innovation'. They reveal that the use of CE on its own does not always lead to positive results and that success in improving innovation capabilities depends on the context, in which CE is applied. Gao, *et al.,* (2000), argues that the extensive applications of computer aided engineering (CAE) technologies are necessary so that the maximum design efficiency and effectiveness can be accomplished prior to initial sample production. The main characteristic of such an approach is depended on the system integration in accordance with the design process.

ETO KEY SEGMENTS

Most of those difficulties are caused by inefficient information; however, the most significant ETO problem is long lead-time. The ETO product delivery lead-time greatly influences the duration of any project. However, the completion of a product is dependent on the timely availability of ETO, Elfving, (2003). Extensive research has been carried out on solutions for reducing the lead-time and implementing time compression in supply chains Tsinopoulos, and Bell, (2008). Lead-time reduction is considered as fundamental for overall business improvement (Womack and Jones, 2003) and a cornerstone for lean thinking. One or a combination of three main strategies - elimination, combination and reduction can be used to reduce the lead-time. Research has identified two types of lead-time reduction in supply chains, reducing the mean lead-time and reducing lead-time variation (Hendericks, *et al. at.*, 2007). According to a study, long lead-time has more than half of its causes associated directly or indirectly with the design phase (Elfving, 2003).

Some of the typical examples are:

- Tiresome compilation and poor consistency of design input
- Modifications due to early commitment and lack of understanding
- Modifications due to design errors
- Primitive practice of supplementary design and endorsements
- Complications requiring a large number of design experts

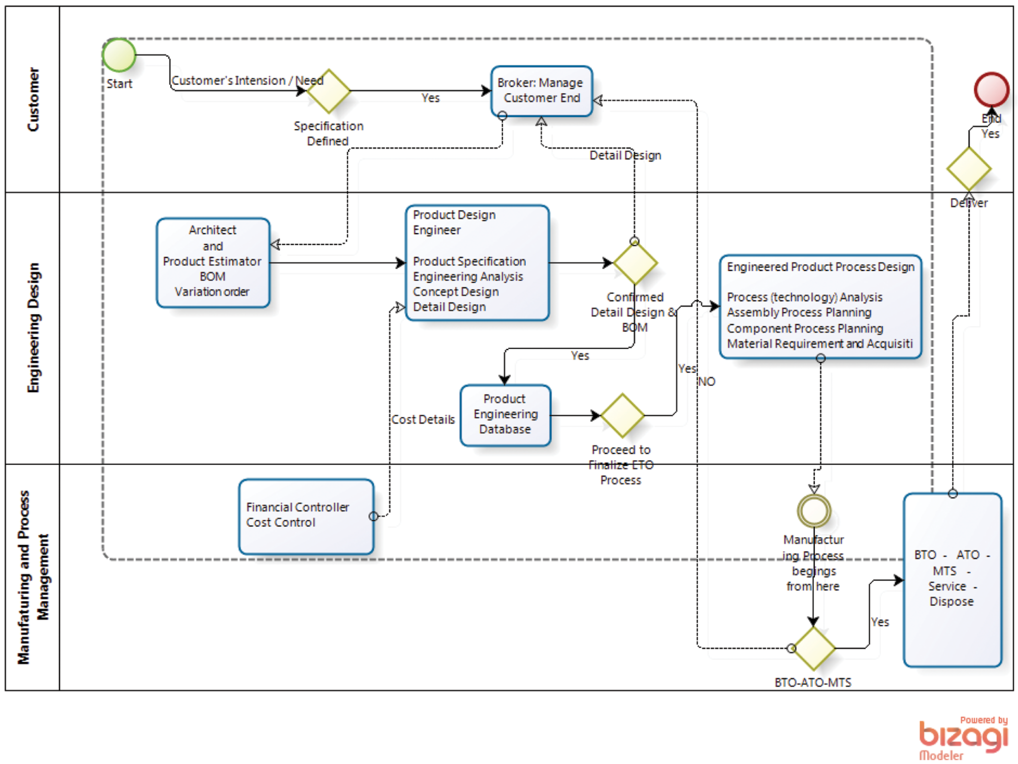
Although there may be many other factors, the causes have their roots partially in poor coordination and poor communication due to the fragmented design environment and the involvement of a large number of various specialists. Therefore, providing a more effective computer-mediated collaborative environment seems to be conducive for resolving the problem.

AN IMPROVED ETO: CE OPERATION PROCESS AND MODEL FOR SCM

ETO is a product development process that starts with a product specification and finishes with an engineering design. In a classical ETO operation, product functionality is the major design focus. Various design testing may be carried out in the design process to verify the product design in progress. The design and standardization of a manufacturing process may be a development focus as well, when intended for serial/volume production.

Figure 2 illustrates a typical Engineer-To-Order (ETO) concurrent engineering approach process model. As already emphasised in the sections above, the whole process always begins with the interest of the customer. The customer's intension or need about a product is passed or received by a broker (Marketer). The confirmed specification in question is then passed on to the product engineering team or partners with the first contact being the architect or the estimator, who then, processes the data to produce a bill of

Figure 2. An improved ETO concurrent engineering (CE) approach - Process model diagram - A typical ETO process model marked in the dotted area: Facilitated by BizAgi process modeller software



materials (BOM). The BOM is then passed on to the product engineer design department. On receiving the product specification which has come to them the design engineer proceeds to perform their engineering analysis as per the product specification to develop the concept design of the product. With the concept design and cost details from the accounts and financial controller's department, a detailed design is then arrived at.

The detailed product design is then sent to a central product engineering database where the engineered process design department accesses it for engineering process analysis (technologies required for production), assembly process planning, and component process planning and finally tested for material requirement and acquisition, then passed on for manufacturing. This is the stage where the Engineer-To-Order process actually finishes and then prepares to enter the manufacturing stages - (Build-To-Order [BTO], Assemble-To-Order [ATO], and Make-To-Stock [MTS]) upon the customer's satisfaction of the ETO product via the broker or marketer. ATO – this is a manufacturing or production approach where all the basic components of the product are already available but not yet assembled, but this is done and sent quickly to the customer once the order for production is received. BTO – this is sometimes also referred to as "Make-To-Order" (MTO). It's a manufacturing or production approach where products are built or made once a confirmed order is received from a customer. This is the current approached employed for highly customized or low quantity production, and it is also the oldest method of order fulfilment in existence. MTO - this is sometimes also referred to as build to stock (BTS). This is an already manufactured and assembled production approach, which depends or is based upon the history or anticipation of supply and demand of the product in question. Thus, large quantity of the product is manufactured, assembled and stocked in warehouses and are supplied when the demand order is received.

These processes flow through from the final delivery of engineered-to-order product that is, if there are no arising faults over time during the ETO process. Finally, disposing off entire if the product exhausts through its life cycle. Thus, ETO is the operation mode that engages in product and process designs. It may include manufacture of the product design if it is one of a kind, but the two (design and manufacture) processes typically occur in sequence. An ETO concurrent engineering approach is an operation in which both design and manufacture processes occur in parallel, in order to optimally minimize the product development time that spans a product specification and delivery of the final product. ETO starts usually working closely with the customers at the stage of specifying a customer need or defining a product specification because when an ETO product is not well defined and confirmed a huge lose will be incurred when the product design progresses. The following section introduces and elaborates on ETO "best practice" as a focal point for this chapter.

ETO BEST PRACTICE

Engineer-to-Order (ETO) is an important production approach in concurrent engineered design and manufacturing; this plays an important role in economies by way of the industrial organizations optimal outputs. Typical ETO products include large electric machine, huge centrifugal pumps, diesel/natural fuel power plant engines, steam turbine, boiler, ship, etc., (Li, 2002). The collaborative design for ETO products has some special engineering design requirements relative to other products, see Box 1:

SUPPLY CHAIN MANAGEMENT TO CRITICAL SUCCESS FACTORS

ETO, a state of the art system tool with collaborative engineering Enterprise Resource Planning (ERP) systems, such as Enterprise Service Architecture (ESA) application processes. Service-Oriented Architecture (SOA) as a platform for this system application and processes provides effective communication for this collaborative engineering system approach, which proves that it is possible to have it all. Thus, clear cost notability and much competence and control among the various partners with a state of the art engineer-to-order ERP software system tools which suits or easily links the design methods of each partner in the design team. The engineer-

Box 1. ETO product design

Multiple objectives: Besides the product structure and its production operation procedures (e.g., routings and routes), ETO product design should also evaluate the feasibility of production planning, services, quality, procurement, etc. That is to say, ETO product design must face to a whole supply chain instead of the independent design activities.
Multiple functions: This is to support the realization of multiple objectives. It is necessary for the department of sales and the ETO enterprise together with its key suppliers and customers; to participate in a collaborative design process instead of only the research and development (R&D) department.
Integrated design: Product design system should not be considered as isolated software but must be closely integrated with other related business systems, e.g., Enterprise Resource Planning (ERP) such as Enterprise Service Architecture (ESA) application processes with Service-Oriented Architecture (SOA) as a platform for applications and processes effective communication, to realize data, functions and process integration (Valle, and Va´zquez-Bustelo, 2009; Campbell, and Mohun, 2007).

ing design software industry is saturated with series of efficient engineering design units that have had to make unnecessary major changes to their software operation processes just to accommodate the reporting requirements of these ERP design software systems. Thus, if the finance and accounting management partners and team's reporting do not adequately reflect the real status of the engineer-to-order design product and their engineering operation process, their final value must be in some doubt.

In an engineer-to-order design environment, estimating the bill of materials (BOM), design process engineering, scheduling and cost are particularly difficult due to time constrains. Hence, identifying the right partners with the requisite capabilities at the right time is very essential for a successful engineer-to-order design process. Since a longer lead-time always incurs extra cost and penalties, engineering product design partner in an engineer-to-order process team need the capabilities to monitor the actual progress against the plan scheduled for the engineer-to-order process basis. These capabilities of the individual engineering product design partners coupled with the collaboration of their various state of the art Enterprise Resource Planning (ERP) engineering design software systems will always achieve a successful engineer-to-order design product for the collaborating partners.

A case example of ETO – CE approach in improving supply chains is the case of an Australian furnishing industry which was subjected to severe competition from global suppliers in the last decade. Supported by the government's Action Agenda, the Furnishing Industry Association of Australia developed the ''Production Efficiency Program'' to improve efficiency of the industry. This "Production Efficiency Program" as attributes and approach similar to that of the ETO – CE approach. Thus, the core of the program was three showcase projects demonstrating how return on investment can be achieved with the application of advanced manufacturing technologies to significantly reduce the design lead-time. Instead of committing large capital expenditure in IT based manufacturing systems, the showcases started by applying concurrent engineering principles by bringing all the design team and organizational partners together on a common platform. Hence, transformed the business practices to adopting latest IT systems for the designing, planning and control of manufacturing supply chain processes. The outcome as reported in the case study of the small furniture company achieved 30% increase in productivity by implementing a new scheduling system that assisted implementation of lean manufacturing and other manufacturing principles and best practices (Mo, 2009).

ETO CLASSIC TOOLS

Classical engineer-to-order state of the art tools for a collaborative engineering design product includes the following listed below in Table 1.

CONCLUSION AND RESEARCH GAPS

ETO in its classical perspective is covered in the previous paragraphs above in a predominantly concurrent or collaborative design engineering process. Due to the need for effective control and reduction of lead-time in the entire ETO supply chain process, it is quite obvious that the design stage is responsible for a chunk of the frequent

Table 1. New and classic ETO tools

New ETO Design Software Tools	Classic ETO Design Software Tools
Team Centre	Solid Works
Siemens NX	Iron CAD
PTC Pro. Engineer	Mechanical CAD
Catia V6	AutoCAD

extended lead-time. Therefore, efficiency in reducing this as well as improving the integration and communication applications and processes in an ETO supply chain is very much an imperative. Thus, for future research, engineering enterprise system tool for engineering Enterprise Resource Planning (ERP) systems such as Enterprise Service Architecture (ESA) application processes with Service-Oriented Architecture (SOA) as a platform for applications and processes will enhance the efficiency required when effectively and efficiently integrated within the ETO supply chain. This approach will also go a long way to reduce the design lead-time and also improving effective communication in the design engineering chain for industrial competitive advantage. As a concluding remark, ETO supply chain managers in manufacturing industries should start to link their operational processes to create responsiveness by advancing product development processes. Employing manufacturing flexibility in a concurrent fashion within ETO supply chain management information systems and information technology systems as well as customer relationship integration earlier in a product development process chain will enhance a competitive advantage for the user.

REFERENCES

Ainscough, M., & Yazdani, B. (2000). Concurrent engineering within British industry. *Concurrent Engineering: Research and Applications*, *8*(1), 2–11.

Balbontin, A. (2000). New product development practices in American and British firms. *Technovation*, *20*, 257–274. doi:10.1016/S0166-4972(99)00136-4

Barba, E. (2001). *Ingeniera Concurrente. Gua parasuImplantacio'n enla Empresa*. Barcelona: Diagnostico y Evaluacion Gestion.

Barber, E. (2009). How to measure the value in the value chains. *International Journal of Physical Distribution & Logistics Management*, *38*(9), 685–698. doi:10.1108/09600030810925971

Briere-Cote, A., Rivest, L., & Desrochers, A. (2010). Adaptive generic product structure modelling for design reuse in engineer-to-order. *Computers in Industry*, *61*, 53–65. doi:10.1016/j.compind.2009.07.005

Calantone, R. J., & Di Benedetto, C. A. (2000). Performance and time to market: Accelerating cycle time with overlapping stages. *IEEE Transactions on Engineering Management*, *47*(2), 232–244. doi:10.1109/17.846790

Cameron, N. S., & Braiden, P. M. (2004). Using business process re-engineering for the development of production efficiency in companies making engineered to order products. *International Production Economics*, *89*, 261–273. doi:10.1016/S0925-5273(02)00448-6

Campbell, S., & Mohun, V. (2007). *Mastering enterprise SOA with SAP Netweaver and mySAP ERP*. Indianapolis, IN: Wiley Publishing, Inc.

Chen, C.-S. (2006). Concurrent engineer-to-order operation in the manufacturing engineering contract industries. *International Journal of Industrial and Systems Engineering*, *1*(1-2), 37–58. doi:10.1504/IJISE.2006.009049

Chopra, S., & Meindl, P. (2001). *Supply chain management, strategy, planning and operation*. Upper Saddle River, NJ: Prentice-Hall.

Combs, L. (2004). The right channel at the right time. *Industrial Management (Des Plaines)*, *46*(4), 8–16.

Conner, M. (2004). The supply chain's role in leveraging PLM. *Supply Chain Management Review*, *8*(2), 36–43.

Cooper, R. G., & Edgett, S. J. (2003). *Best practices in product innovation: What distinguishes top performers*. Ancaster, Canada: Product Development Institute Inc.

Demeter, K., & Mateusz, Z. (2011). The impact of lean practices on inventory turnover. *International Journal of Production Economics*, *133*, 154–163. doi:10.1016/j.ijpe.2009.10.031

Elfving, J. (2003). *Exploration of opportunity to reduce lead time for engineered-to-order products*. PhD Dissertation, University of California, Berkeley.

Gao, J., X., *et al.* (2000). Implementation of concurrent engineering in the suppliers to the automotive industry. *Journal of Materials Processing Technology*, *107*, 201–208. doi:10.1016/S0924-0136(00)00669-5

Gosling, J., & Naim, M. M. (2009). Engineer-to-order supply chain management: A literature review and research agenda. *International Journal of Production Economics*, *122*, 741–754. doi:10.1016/j.ijpe.2009.07.002

Gunasekaran, A., Lai, K. C., & Cheng, T. C. (2008). Responsive supply chain: A competitive strategy in networked company. *International Journal of Management Science*, *36*, 549–564.

Hendericks, K. B. (2007). The impact of enterprise systems on corporate performance: A study of ERP, SCM and CRM system implementations. *Journal of Operations Management*, *25*(1), 65–82. doi:10.1016/j.jom.2006.02.002

Herder, P. M., & Weijnen, M. P. C. (2000). A concurrent engineering approach to chemical process design. *International Journal of Production Economics*, *64*, 311–318. doi:10.1016/S0925-5273(99)00068-7

Holweg, M. (2005). The three dimensions of responsiveness. *International Journal of Operations & Production Management*, *25*(7), 603–622. doi:10.1108/01443570510605063

Ketchen, D. J., Hult, T. M., Rebarick, W., & Meyer, D. (2008). Best value supply chains: A key competitive weapon for the 21st century. *Business Horizons*, *51*, 235–243. doi:10.1016/j.bushor.2008.01.012

Kim, B. (2005). *Mastering business in Asia: Supply chain management*. Singapore: Wiley.

Koufteros, X. (2001). Concurrent engineering and its consequences. *Journal of Operations Management*, *19*, 97–115. doi:10.1016/S0272-6963(00)00048-6

Kumar, A., Fantazy, K. A., & Kumar, U. (2006). Implementation and management framework for supply chain flexibility. *Journal of Enterprise Information*, *19*(3), 303–319. doi:10.1108/17410390610658487

Lee, H. L. (2004). The triple-A supply chain. *Harvard Business Review*, *82*(10), 102–112.

Li, S. H., Rao, S. S., Nathan, R. T., & Nathan, B. R. (2006). The impact of supply chain management practices on competitive advantage and organizational performance. *Omega*, *34*, 107–124. doi:10.1016/j.omega.2004.08.002

Li, X. P. (2002). *Optimization methods of multilevel hierarchical planning architecture of engineer-to-order enterprises*. PhD Dissertation, Harbin Institute of Technology, Harbin, China.

MEEP. (2009). *Concurrent engineering materials.* The Manufacturing Education Partnership. Penn State University of Puerto Rico-Mayaguez & University of Washington Sandia National Lab.

Mo, J. P. T. (2009). The role of lean in the application of information technology to manufacturing. *Computers in Industry*, *60*, 266–276. doi:10.1016/j.compind.2009.01.002

O'Marah, K. (2003). The business case for PLM. *Supply Chain Management Review*, *7*(6), 16–18.

O'Marah, K., & Hofman, D. (2010). *The AMR supply chain top 25 for 2010*. Stamford, CT: Gartner Research Publication.

Pandit, A., & Yimin, Z. (2007). An ontology based approach to support decision-making for the design of ETO (Engineer-To-Order) products. *Automation in Construction*, *16*, 759–770. doi:10.1016/j.autcon.2007.02.003

Patel, A. (2008). *Best practices in front-end design*. A Bentley White Paper, design and FEED applications.

Porter, M. E. (1990). The competitive advantage of nation. *Harvard Business Review*, *68*(2), 73–93.

Portioli-Staudacher, A., *et al.* (2003). *Implementation of concurrent engineering: A survey in Italy and Belguim.*

Reichhart, A., & Holweg, M. (2007). Creating the customer-responsive supply chain: A reconciliation of concepts. *International Journal of Operations & Production Management*, *27*(11), 1144–1172. doi:10.1108/01443570710830575

Riezebos, J., Klingenberg, W., & Hicks, C. (2009). Lean production and Information Technology: Connection or contradiction? *Computers in Industry*, *60*, 237–247. doi:10.1016/j.compind.2009.01.004

SAP AG. (2003). *Manufacturing strategy: An adaptive perspective.* SAP White Paper by SAP SCM.

Swafford, P., Gosh, S., & Murthy, N. (2006). A framework for assessing value chain agility. *International Journal of Operations & Production Management*, *26*(2), 118–140. doi:10.1108/01443570610641639

Tennant, C., & Roberts, P. (2000). A faster way to create better quality products. *International Journal of Project Management*, *19*, 353–362. doi:10.1016/S0263-7863(00)00010-7

Teresko, J. (2004). The PLM revolution. *Industry Week*, *253*(2), 32–36.

Tsinopoulos, K., & Bell, K. (2008). Supply chain integration systems by small engineer to order companies: The challenge of implementation. *Journal of Manufacturing Technology Management*, *21*(1), 50–62. doi:10.1108/17410381011011489

Umemoto, K., Endo, A., & Machaco, M. (2004). From Sashimi to Zen-in. The evolution of concurrent engineering at Fuji Xerox. *Journal of Knowledge Management*, *8*(4), 89–99. doi:10.1108/13673270410548504

Valle, S., & Vazquez-Bustelo, D. (2009). Concurrent engineering performance: Incremental versus radical innovation. *International Journal of Production Economics*, *119*, 136–148. doi:10.1016/j.ijpe.2009.02.002

Wang, Z., Zhan, D., & Xu, X. (2006). Service-oriented infrastructure for collaborative product design in ETO enterprises. *Proceedings of 10th International Conference on Computer Supported Cooperative Work in Design.*

Womack, J., & Jones, D. T. (2003). *Lean thinking: Banish waste and create wealth in your corporation* (2nd ed., pp. 15–36). New York, NY: Free Press.

Yassine, A., & Braha, D. (2003). Four complex concurrent engineering and the design structure matrix method. *Concurrent Engineering: Research and Applications*, *11*(3), 165–176. doi:10.1177/106329303034503

Zhou, H., & Benton, W. C. (2007). Supply chain practice and information sharing. *Journal of Operations Management*, *25*, 1348–1365. doi:10.1016/j.jom.2007.01.009

KEY TERMS AND DEFINITIONS

Concurrent Engineering: Concurrent engineering in this sense is moderate collaborative engineering approach that is much more focused on the engineering product design process of concurrent cross-functional partner organizations together in order to create products that are much better, much affordable and with a much reduced lead-time.

Customer Oriented Supply Chains Management: This is where the entire supply chain management is geared or focused towards cordinating events for the full satisfaction of the end customer.

Engineer-to-Order: ETO is a product development process, which starts with a product specification and finishes with an engineering design as its deliverable.

ETO-CE Best Practice Approach: The ETO-CE operational process is termed as the "best practice" approach in this chapter.

ETO-CE Operational Process Model: An ETO process flow model which utilizes the concurrent engineering operational principle in improving ETO supply chains.

ETO-CE SCM integration process: This is where some ERP / IT systems are integrated in the ETO-CE supply chains to enhance effective and efficient improvement in the entire ETO supply chain management.

Supply chain: a system of technology, activities, information, people as well as organizations and resources involved in moving a product or service from supplier to customer at the right place and at the right time.

Supply Chain Management: Supply Chain Management is the management of the entire interconnected chain of an organization including the activities, technology, information, the people as well as the all the other resources involved.

APPENDIX

ATO	Assemble To Order
BOM	Bill Of Materials
BTO	Build To Order
CAD	Computer Aided Design
CAE	Computer Aided Engineering
CAM	Computer Aided Manufacture
CE	Concurrent Engineering
ERP	Enterprise Resource Planning
ESA	Enterprise Service Architecture
ETO	Engineer To Order
IDA	Institute for Defense Analysis
IPD	Integrated Product development
JIT	Just In Time
LM	Lean Manufacture
MTS	Make To Stock
NPI/D	New Product Introduction / Development
PLC	Product Life Cycle
R&D	Research & Development
SC	Supply Chain
SCM	Supply Chain Management
SOA	Service Oriented Architecture
XML	Extensible Mark-up Language

This work was previously published in Customer-Oriented Global Supply Chains: Concepts for Effective Management, edited by Ephrem Eyob and Edem G. Tetteh, pp. 112-128, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Section 8
Emerging Trends

*This section highlights research potential within the field of Industrial Engineering while exploring uncharted areas of study for the advancement of the discipline. Introducing this section are chapters that set the stage for future research directions and topical suggestions for continued debate, centering on the new venues and forums for discussion. A pair of chapters on supply chain management and green computing makes up the middle of the section of the final 14 chapters, and the book concludes with a look ahead into the future of the Industrial Engineering field, with "*Zero-Downtime Reconfiguration of Distributed Control Logic in Industrial Automation and Control.*" In all, this text will serve as a vital resource to practitioners and academics interested in the best practices and applications of the burgeoning field of Industrial Engineering.*

Chapter 96
Advanced Technologies for Transient Faults Detection and Compensation

Matteo Sonza Reorda
Politecnico di Torino, Italy

Luca Sterpone
Politecnico di Torino, Italy

Massimo Violante
Politecnico di Torino, Italy

ABSTRACT

Transient faults became an increasing issue in the past few years as smaller geometries of newer, highly miniaturized, silicon manufacturing technologies brought to the mass-market failure mechanisms traditionally bound to niche markets as electronic equipments for avionic, space or nuclear applications. This chapter presents the origin of transient faults, it discusses the propagation mechanism, it outlines models devised to represent them and finally it discusses the state-of-the-art design techniques that can be used to detect and correct transient faults. The concepts of hardware, data and time redundancy are presented, and their implementations to cope with transient faults affecting storage elements, combinational logic and IP-cores (e.g., processor cores) typically found in a System-on-Chip are discussed.

1. INTRODUCTION

Advanced semiconductor technologies developed in the past few years are allowing giant leaps forward to the electronic industry. Nowadays, portable devices are available that provide several orders of magnitude more computing power than top-of-the-line workstations of few years ago.

DOI: 10.4018/978-1-4666-1945-6.ch096

Advanced semiconductor technologies are able to achieve such improvements by shrinking the feature size that is now at 22 nm and below, allowing integrating millions of devices on a single chip. As a result, it is now possible to manufacture an entire system (encompassing processors, companion chips, memories and input/output modules) on a single chip. Smaller transistors are also able to switch faster, thus allowing operational frequencies in the GHz range. Finally, low operational voltages are possible, significantly reducing the energy needs of complex chips.

All these benefits have however a downside in the higher sensitivity of newer devices to soft errors. The reduced amount of charge needed to store memory bits, the increased operational frequencies, as well as the reduced noise margins coming from lower operational voltages are making the occurrence of soft errors, i.e., unexpected random failures of the system, more probable during system lifetime.

Among the different sources of soft errors, radiation induced events are becoming more and more important, and interest is growing on this topic from both the academic and the industrial communities.

As described in (Dodd et al., 2004), when ionizing radiations (heavy ions or, protons in space, neutrons, and alpha particles in the earth atmosphere) hit the sensitive volume of a semiconductor device (its reserve biased depletion region) the injected charge is accelerated by an electric field, resulting in a parasitic current than can produce a number of effects, generally referred to as Single Event Effects (SEEs). Single Event Latchup (SEL) is the destructive event that takes place when the parasitic current triggers non-functional structures hidden in the semiconductor device (like parasitic transistors that shorten ground lines to power lines, which should never conduct when the device is operating correctly). Single Event Upset (SEU) is the not-destructive event that takes place when the parasitic current is able to trigger the modification of a storage cell, whose content flips from 0 to 1, or vice-versa. In case the injected charge reaches the sensitive volume of more than one memory device, multiple SEUs may happen simultaneously, causing the phenomenon known as Multiple Bit Upset (MBU). Finally, Single Event Transient (SET) is the not-destructive event that takes place when the parasitic current produces glitches on the values of nets in the circuit compatible with the noise margins of the technology, thus result in the temporary modification of the value of the nets from 0 to 1, or vice-versa.

Among SEEs, SEL is the most worrisome, as it corresponds to the destruction of the device, and hence it is normally solved by means of SEL-aware layout of silicon cells, or by current sensing and limiting circuits. SEUs, MBUs, and SETs can be tackled in different ways, depending on the market the application aims at. When vertical, high-budget, applications are considered, like for example electronic devices for telecom satellites, SEE-immune manufacturing technologies can be adopted, which are by-construction immune to SEUs, MBUs, and SETs, but whose costs are prohibitive for any other market. When budget-constrained applications are considered, from electronic devices for space exploration missions to automotive and commodity applications, SEUs, MBUs and SETs should be tackled by adopting fault detection and compensation techniques that allow developing dependable systems (i.e., where SEE effects produce negligible impacts on the application end user) on top of intrinsically not dependable technologies (i.e., which can be subject to SEUs, MBUs, and SETs), whose manufacturing costs are affordable.

Different types of fault detection and compensation techniques have been developed in the past years, which are based on the well-known concepts of resource, information or time redundancy (Pradhan, 1996).

In this chapter we first look at the source of soft errors, by presenting some background on radioactive environments, and then discussing how soft errors can be seen at the device level. When then present the most interesting mitigation techniques organized as a function of the component they aims at: processor, memory module, and random logic. Finally, we draw some conclusions.

2. BACKGROUND

The purpose of this section is to present an overview of the radioactive environments, to introduce the reader to the physical roots of soft errors. Afterwards, SEEs resulting from the interaction of ionizing radiation with the sensitive volume of semiconductor devices are discussed at the device level, defining some fault models useful to present fault detection and compensation techniques.

2.1. Radioactive Environments

The sources of radiations can be classified in different ways, depending on where the system is deployed. We can consider three so-called radiation environments: space, atmospheric and ground radiation environments (Barth et al., 2003).

The space radiation environment is composed of particles trapped by planetary magnetospheres (protons, electrons, and heavier ions), galactic cosmic ray particles (heavy ions and protons) and particles from solar events, such as coronal mass ejection and flares, which produce energetic protons, alpha particles, heavy ions, and electrons (Barth et al., 2003). The maximum energy the particles have ranges from 10 MeV for trapped electrons up to 1 TeV for galactic cosmic rays (1 eV being equivalent to $16x10^{-21}$ Joules). Due to the very high energies involved, shielding may not be effective in protecting circuits, and therefore the impact of ionizing radiation on electronic devices should be investigated deeply, to devise effective fault compensation techniques.

Atmospheric and ground radiation environments are quite different with respect to the space environment. Indeed, when cosmic ray and solar particles enter the Earth's atmosphere, they interact with atoms of nitrogen and oxygen, and are they are attenuated. The product of the attenuation process is a shower of protons, electrons, neutrons, heavy ions, muons, and pions. Among these particles, the most important ones are neutrons, which start to appear from 330 Km of altitude. Neutron density increases up to the peak density found at about 20 Km of altitude, and then it decreases until the ground level, where the neutron density is about 1/500 of the peak one (Taber et al., 1995). The maximum energy observed for the particles in the atmospheric radiation environment is about some hundreds of MeV.

At the ground level, beside neutrons resulting from the interaction of galactic cosmic ray and sun particles with the atmosphere, second most important radiation source is the man-produce radiation (nuclear facilities).

No matter the radiation environment where the system is deployed, we have that when radiations interact with semiconductor devices two types of interactions can be observed: atomic displacement or ionization. Atomic displacement corresponds to modifications to the structure of silicon device, which may show for example displaced atoms, and it is out of the scope of this chapter. Conversely, the latter corresponds to the deposition of energy in the semiconductor, and it is focused in this chapter.

Radiations may inject charge (i.e., ionize) a semiconductor device in two different ways: direct ionization by the particle that strikes the silicon, or ionization by secondary particles created by nuclear reactions between the incident particle and the silicon. Both methods are critical, since both of them may produce malfunctions (Dodd et al., 2003).

When an energetic particle passes through a semiconductor material it frees electron-hole pairs along its path, and it loses energy. When all its energy is lost, the particle rests in the semiconductor, after having travelled a path length called particle range. The energy loss per unit path length of a particle travelling in a material is known as linear energy transfer (LET), measured in $MeVcm^2/mg$: the energy loss per unit path length (MeV/cm) divided by the material density (mg/cm^3). As an example, a particle having an LET of 97 $MeVcm^2/mg$ deposits a charge of 1 pC/mm in Silicon.

Heavy ions inject charges in a semiconductor device by means of the mechanism called direct ionization (Dodd et al., 2003). Protons and neutrons do not produce enough charge by direct ionization to cause single-event effects, although recent studies showed that single-event effects due to direct ionization by means of protons are possible (Barak et al., 1996) in highly scaled devices. Indirect ionization is the mechanism through which protons and neutrons produce single-event effects. Proton, or neutron, entering a semiconductor device produces atomic reactions with silicon atoms, originating by-products like alpha or gamma particles. These by-products can deposit energy along their paths by direct ionization, causing single-event effects (Dodd et al., 2003).

2.2. A Device-Level View of Radiation Effects

The parasitic current induced by (direct or indirect) ionization can result in a number of different device-level effects, depending on when and where the charge injection takes place. We can broadly classify the device-level effects as destructive and not destructive. As far as digital electronic devices are considered, the most important destructive SEE is the SEL, while the most relevant not destructive SEEs are SEUs/MBUs and SETs. The following sub-sections describe these phenomena.

2.2.1. Single Event Latchup

Semiconductor devices like pMOS or nMOS transistor contains parasitic structures, composed of two bipolar transistors forming a silicon-controlled rectifier, as depicted in Figure 1.

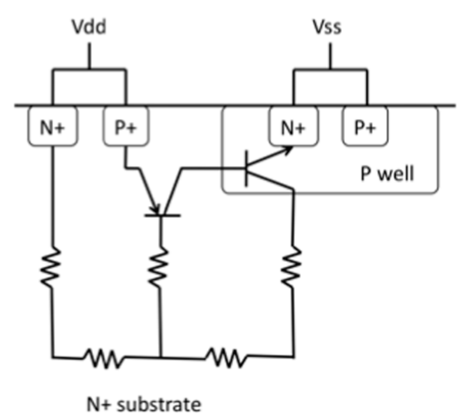
If the current resulting from ionization triggers the parasitic structure, a short circuit between power and ground lines is activated, resulting in a high current flowing in the device. In case such a current is not stopped promptly, permanent damage of the device is likely to happen.

2.2.2. Single Event Upset

As explained in (Dodd et al., 2003), DRAM technology refers to those devices that store bits as charge in a capacitor. In these devices no active information regeneration exists; therefore, any disturbance of the stored information provoked by ionizing radiations is persistent until it is corrected by a new write operation. Any degeneration of the stored charge that corresponds to a signal level outside the noise margin of the read circuit is sufficient to provoke an error. The noise margin is related to the memory critical charge, Q_{crit}, which is defined as the minimum amount of charge collected at a sensitive node that is necessary to cause the memory to change its state.

The most important SEU source in DRAMs is the SEE charge collection within each capacitor used to store bits in the DRAM. These errors are caused by a single-event strike in or near either the storage capacitor or the source of the access transistor. Such a strike affects the stored charge by the collection of induced charge (Dodd et al., 2003). Such error corresponds normally to a transition of the stored bit from 1 to 0 (May et al., 1979). However, the ALPEN effect (Rajeevakumar et al., 1988) makes transitions from 0 to 1 as well. SEUs can also occur in DRAMs due

Figure 1. The parasitic silicon-controller rectifies in a nMOS device



to charge injection in the bit lines (Dodd et al., 2003), or a combination of charge injection close to the bit capacitor and the bit line (Rajeevakumar et al., 1988).

SEUs are originated in SRAMs in according to a different phenomenon with respect to DRAMs. Indeed, in SRAM the information bit in restored continuously, by means of a two inverters forming a feedback loop. When ionizing radiation injects charge into a sensitive location in a SRAM, a transient current is originated in the affected transistor. The existing feedback counterbalances the injected current, trying restoring a correct (stable) configuration. In case the current originated by the injected charge is higher than the restoring current, a voltage pulse occurs that lead to the corruption of the stored bit (Dodd et al., 2003). In case charge injection affects multiple memory cells, the resulting effect correspond to multiple SEUs happening simultaneously; such event is known as Multiple Cell Upset.

2.2.3. Single Event Transient

The progressive decreasing of the minimum dimensions of integrated circuits, accompanied by increasing operating frequencies lead on the one side the possibility of using lower supply voltages with very low noise margins but on the other side it make integrated circuits (ICs) more sensitive to Single Event Transient (SET) pulses (Baumann, 2005). In details, the shrinking technology process decreases the manufacturing sizes reducing the charge required to identify a logic state. The result is that the parasitic current provoked by the ionizing radiation is inducing a pulse effect, also called SET.

The high-energy recoil and the proton-induced nuclear reactions are behind the SET generation mechanisms. In details, low-angle protons as well as heavy ions affect the silicon area close to a junction resulting in energy-loss of the signal and thus in observable SETs. The shape of SETs may be different depending on the source of the charge deposition and the conditions related to the tracks the considered high energy particle. More is the charge, more the SETs is characterized by peak heights and widths dependent from the space-charge effects (i.e. the heavy-ions generate this kind of effect).

The high-injection of charge provokes a variation of the external fields internally to the considered region. The carriers within the considered region are drifted by the ambipolar diffusion reaching the edge of the plasma edge, where they drift to the external region filed thus provoking a pulse of current (i.e. The SET effect). The pulse current is drastically smaller than the normal current induced by a logic-drift. Depending on the carrier levels and on the threshold level, the penetration within the external field may change thus resulting in a different SET width.

This phenomenon has been deeply investigated by experimental analysis where researchers control the radiation beam spot size, the intensity and the positions thus changing the charge injection of the particle strikes. The analysis of SET pulses has been obtained collecting datasets from pulse monitors or pulse shape analysis (PSA) that allow to generate the images of the induced SET measuring the charge and the maximum current, extracted over an area of interest.

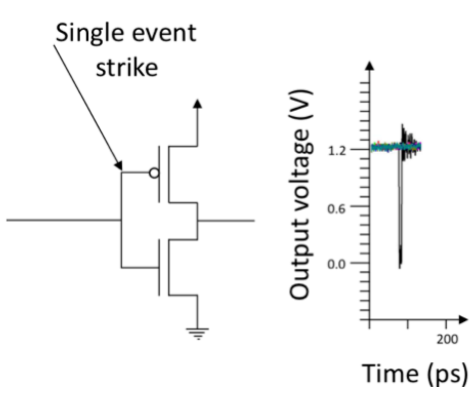
Many works investigated the nature of these events, measuring the propagation of the transient pulses through the combinational logic and routing resources in specifically designed circuits. These circuits have been demonstrated to be an effective solution for study SET effects in logic and routing resources of ICs ranging from ASICs to regular fabric such as FPGAs. In particular, for CMOS-based devices, energetic particle strikes can cause a transient voltage pulse that propagates through the circuits and may become an uncorrected data in case they are stored by a sequential element, causing disruption of the expected circuit operations. The upset rates due to SETs are dependent on the pulse width of the SET and the clock frequency of the circuit under analysis; in particular, with higher clock frequencies there are more latching clock edges to capture SET.

In particular, the SET pulse is originated by the energy loss released in the silicon structure characterized by the nMOS and pMOS transistors, as illustrated in the principal mechanism in Figure 2. The energy distribution is spatially uniform. This is due to the spatially non-variant loss of energy, thus it is assumed that the maximal peak of current as well as the rise and fall time are predominant. This is due to the electromagnetic field generation with runs with a specified angle with respect to the metallization.

The radiation environment, such as the one for space and avionic applications, is extremely stressing digital circuits for the generation and propagation of SET pulses, which have become a growing concern.

Depending on which is the affected circuit location, SETs may propagate through circuits and generate errors in digital combinational logic gates or a memory cells. Large distributions of SETs have been measured along combinational logic chains irradiated with heavy ions. The results reported large SET widths, up to 2 ns for simpler inverter chains. Several laser experiments demonstrated that SETs measured at the output of a logic gate chain undergo to pulse broadening effect induced by the propagation of SETs in long chains of identical inverters (Ferlet-Cavrois, 2007). This effect has been defined as Propagation-Induced-Pulse-Broadening (PIPB). It has been demonstrated that the SET pulse broadening is induced by the variation of the body potential, depending on the previous states of the transistor, this effect is related to the charge and discharge phenomenon of the transistor body that modifies the electrical response of the transistor to the current pulse propagation. Besides radiation experiments analysis, several other works investigated the nature of these events studying the propagation of the transients through the combinational logic and routing resources of ad-hoc designed circuits. These circuits have been demonstrated to be an efficient way of studying SET effects on logic and routing resources; however, they are extremely far from being representative of realistic designs. In order to study SET effects on real designs, three kinds of analysis approaches have been adopted. First of all, radiation experiments on complex circuits and real applications provide the behavior of designs implemented in such technology under radiation. In this case, the more is the desired accuracy to study precise effects, the more it is necessary to instrument and change the circuit. Secondly, laser techniques can be used to inject pulses in a desired point of the circuit at a precise energy and controlled pulse width. On the other hand, laser testing is an extremely accurate technique that can be used once the pulse lasers are focused on the active area and the energy is adjusted at the right threshold level. The original pulse width and amplitude generated either by radiation beam or laser test are almost impossible to be directly measured, therefore the two techniques cannot be directly used to precisely characterize the SET propagation effects. Between these two techniques, electrical pulse injection can be applied to the device under analysis with known pulse signal properties (width and amplitude). In case the electrical injection is performed externally to the integrated circuit, it requires a

Figure 2. A SET generation mechanisms on an inverter gate and the measured SET pulse output voltage



high-performance signal generator capable to provide appropriate SET pulses (shorter than 1 ns). However, this technique requires advanced characterization of the integrated circuits pads, since the injected pulse is applied to the external pad, the pulse shape is subjected to distortion effect due to the input buffer and the internal interconnections that stem to the injection location. Thus the shape of the injected SET may be modified before reaching the desired injection point.

The analysis techniques mentioned above allowed to determine that SETs pulse width and amplitude are intrinsically related to the numbers of logic gates traversed and the capacitive load of each logic gate. Researchers demonstrated that SET may be broadened or compressed depending on these parameters, and in particular the capacitive load plays an imperative role in the pulse width of the generated SET (Sterpone, 2009) (Kastensmidt Lima, 2008).

3. MITIGATION TECHNIQUES

Designers facing the problem of developing dependable systems (i.e., systems that operate without harming its users even in present on SEE-induced errors) need to adopt fault detection and compensation techniques. Indeed, as a matter of fact mainstream semiconductor technologies are prone to SEE, and radiation-hardened technologies insensitive to SEEs are too expensive for being widely adopted.

3.1. Detecting and Compensating SEL

SEL can be solved using two alternative approaches, one working at system level, and a second one working at layout level.

The first one uses current sensors at the system level to detect the excessive current flowing when the SEL is triggered to promptly shutdown the power supply. As the device is powered off, the SEL is stopped, and after a rest period sufficient for the injected charge to annihilate, the device can be power on again. The main drawback of this approach is that the circuit state is lost at power off and therefore a suitable synchronization phase is needed when the device is power on again.

The second approach uses SEL-aware design guidelines during the layout of transistors. By exploiting guard rings, the parasitic silicon controller rectifier is broken, so that SEL are likely not to take place (Troutman, 1983). To further improve such layout-based techniques, special manufacturing process can be exploited. To further reduce the SEL occurrence probability silicon grown on epitaxial layer, or silicon on insulator, is preferable over bulk silicon.

3.2. Detecting and Compensating SEUs, MCUs and SETs

This section presents techniques developed for detecting and compensating not-destructive SEEs. Several approaches are presented tackling SEUs, MCUs, and SETs; for the sake of clarity, the approaches are organized as a function of the component they aim at: processor, memory or random logic.

The presented approaches decline in different ways the well-known concepts of hardware, information and time redundancy (Pradhan, 1996).), i.e., the use of more hardware, stored data or computing time than actually needed for implementing the functionalities the system has to provide, to guarantee that the system operates correctly (or at least safely, i.e., without damaging the system user) even in presence of errors provoked by faulty system components

3.2.1. Processors

In this section we focus only on transient faults that may affect the processor the system includes. We thus assume that memories and other chips (e.g., input/output modules) located outside the processor chip are fault free.

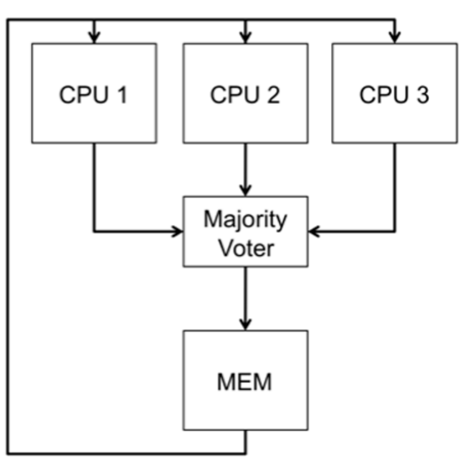
As far as techniques for detecting and compensating transient faults in processors are considered, we can classify them in two broad categories: techniques based on processor redundancy, which achieve fault detection and compensation using more than one processor chip, and techniques based on time and/or information redundancy, which achieve fault detection and compensation using one processor chip in combination with specially crafted software.

As far as techniques exploiting processor redundancy are considered we can further classify them in techniques exploiting passive redundancy, and techniques adopting active redundancy (Pradhan, 1996).

3.2.1.1. Passive Redundancy

The concept of *passive redundancy* consists in replicating the processor core three times, and in equipping the obtained processor module with a majority voter, according to the architecture depicted in Figure 3 that is known as Triple Module Redundancy (TMR) For the sake of simplicity we showed here only the interface between the processor and the memory. In general any bus exiting the processor has to be replicated three times and voted.

Figure 3. Passive redundancy architecture



The three processors of the TMR architecture are synchronized and execute the same software. According to this principle, every write operation toward the memory is executed three times; therefore, under the single-fault assumption, the majority voter (which is supposed to be fault tree) is able to decide correctly the address and the value of the memory involved in every write operation.

The most interesting property of the passive redundancy is the capability of providing a correct service even in presence on one error. Indeed, even in case one of the processor is behaving erroneously, the majority voter is able to decide the correct output that has to be sent to the memory. This property is known as fault masking, because every single fault affecting one of the processors will never have the chance to propagate to the voter output (Pradhan, 1996).

Although very effective in compensating the occurrence of transient faults, the passive redundancy architecture as presented so far does not produce any mechanism to remove the fault from the faulty processor. Considering the architecture of Figure 3, let us suppose that the tree processors are hit by a flux of ionizing radiation, and that CPU 2 experiences one SEU in its program counter. Since the modification of the CPU 2 program counter, the three processors are no longer synchronized. Indeed, CPU 1 and CPU 3 continue to execute the correct program sequence, while CPU 2 executes a different one. However, the system continues to operate correctly, as the majority voter is still able to decide the correct output using the values coming from the two fault-free processors. As the flux of ionizing radiations continues to hit the processors, it is possible that a new SEU hits either CPU 1 or CPU 3. In this event, the majority voter is no longer able to decide the correct output because two processors out of three are faulty. As result, passive redundancy should be enriched with suitable techniques to remove transient faults after their detection, to guarantee that the three processors are always synchronized.

The general algorithm for fault removal encompasses the following steps that perform processor synchronization:

1. Upon fault detection, the normal operations are stopped, and the three processors are forced to execute a procedure to save in memory their context (i.e., the content of the registers, program counter, and program status word). As the values are sent to the memory via the majority voter, a correct context will be stored.
2. The three processors are reset, and possibly powered off and on. As a result, all the memory elements are initialized to a correct value, removing any transient fault.
3. The correct context is copied from the memory to the three processors. At the end of this step, the three processors are synchronized, and the normal operations can be resumed.

Processor synchronization is a complex task that may be simplified by taking advantage of the peculiarity of the application where the system is employed in. For example, in case of cyclic tasks, processor synchronization can takes place at the end of each cycle by resetting the processors, and avoiding context savings.

During processor synchronization, the system is unavailable to the user, as normal operations are stopped. In case outages of service are not allowed, a more complex architecture based on N Module Redundancy, with N > 3, should be used.

Passive redundancy enriched with processor synchronization is currently used in a high-end processing module: the single-board computer for space, scs750, developed and commercialized by Maxwell technologies (http://www.maxwell.com/microelectronics/products/sbc/scs750.asp). The scs750 employs three PowerPC 750FX processors working synchronously, implementing the TMR architecture. The majority voter is implementing into a radiation-hardened FPGA that, being immune to radiations, guarantees fault-free operations. In case an error is detected, processor synchronization is performed in less then 1 msec.

3.2.1.2. Active Redundancy

The concept of *active redundancy* consists in tackling with faults in two consecutive steps: fault detection, followed by fault correction.

Fault detection, as the name implies, is the step during which the behavior of a system is monitored looking for symptoms of errors. While the system behaves correctly, the outputs it provides are forwarded to the user. As soon as, a fault is detected, a safe output is sent to the user, and the fault correction step is initiated. Safe outputs can be seen as output values not expected during correct operations, which are harmless to the user. As an example, a system designed according to the active redundancy concept may provide functional outputs, along with an error alarm output. The user of the system relies on functional outputs as long as the error alarm is not active. Whenever the error alarm is active, functional outputs are discarded until the error alarm is deactivated. Fault detection can be implemented in two ways:

- Instruction-level fault detection consists in monitoring each read/write operation the processor performs toward the external memory.
- Task-level fault detection consists in monitoring the values written in memory as soon as the processor completes the execution of a task (e.g., a sequence of instructions).

Fault correction is the step during which the fault is removed from the system, and its correct operations are restored. In case of transient faults, this may correspond to processor reset to bring all the memory elements to a known-good initial value.

As an example of active redundancy implementing an instruction-level fault detection scheme, we can consider the lockstep approach enriched with checkpoint and rollback recovery we introduced in (Abate et al., 2008).

Aiming at detecting errors affecting the operation of the processor, the lockstep technique uses two identical processors running in parallel the same application. The processors are synchronized to start from the same state. They receive the same inputs, and therefore they should evolve among the same states at every clock cycle, unless an abnormal condition occurs. This characteristic allows for the detection of errors affecting one of the processors through the periodical comparison of the processors states. Comparison of the processor states, called consistency check, is performed after the program executed for a predefined amount of time or whenever an activation condition is met (e.g., a value is ready for being committed to the program user or for being written in memory). In case of mismatch, the execution of the application is interrupted, and the processors must resume from an error-free state.

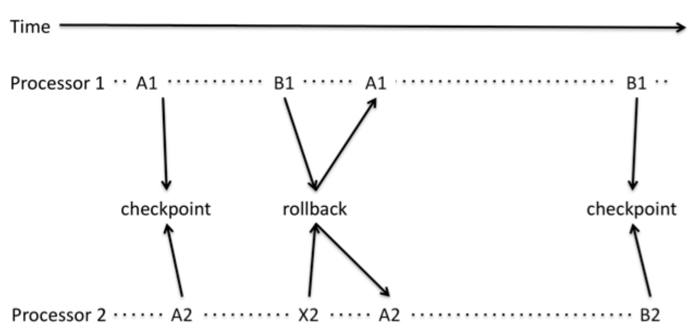
Checkpoints are used to keep a backup copy of the last error-free state in a safe storage. Whenever a consistency check signals a match among processors state, a copy of all information required to restore the processors to that state is saved in a storage device insensitive to ionizing radiation. This set of information is named context, and defines univocally the state of the processor (it can include the content of the processor's registers, the program counter, the cache, the main memory, etc.).

If the consistency check fails, rollback is performed to return both processors to the previous error-free state. The processors are reset, and the previously saved context is copied into them.

Figure 4 shows an example of application execution flow using lockstep coupled with checkpoint and rollback recovery. The arrow on the left indicates the timeline (T).

Initially, processor 1 executes the application until it reaches a predefined point. The context of processor 1 at this point is A1. Then, processor 2 executes the same application, reaching the same point with context A2. When both processors reached the same point, their contexts are compared and if they match a checkpoint is executed. Then, the execution of the application is resumed, with processor 1 performing another portion of the application until it reaches a second point, with context B1, and then processor 2 executes the same portion of the application, stopping at the same second point, with context B2. At this point a new consistency check is executed and, if no error is detected, a new checkpoint is performed, saving contexts B1 and B2, and so on, until the whole application has been successfully executed by both processors.

Figure 4. Example of execution of rollback recovery using checkpoint



Now, let us suppose that, as shown in Figure 4, one SEU occurs and causes one error while processor 2 is processing the second portion of the application. When it reaches the second predefined point and the consistency check is performed, the state of processor 2 is X2, instead of B2, which indicates that one error occurred and that a rollback is needed.

Rollback restores both processors to their last error-free states using the context saved during the last checkpoint, i.e., contexts A1 and A2. The execution of the application is then resumed as previously described, with processor 1 and then processor 2 executing the same portion of the application, and if no other error occurs the processors finally reach the correct states B1 and B2 and a new consistency check is performed, saving contexts B1 and B2. As a result, the error caused by the SEU was during consistency check, it is removed thanks to rollback, and it is corrected by repeating the execution application.

A particularly critical aspect in lockstep is the criteria used to define under which condition the application should be interrupted and a consistency check performed, as it can impact the performance of the system, the error detection latency, and the time required to recover from an error. Checking and saving processors context at every clock cycle provides the shortest fault detection and error recovery times. However, it entails unacceptable performance penalties. Conversely, long intervals between consecutive checkpoints may lead to catastrophic consequences, as the error is free to propagate through the system. Therefore, a suitable trade-off between the frequency of checkpoints, error detection latency and recovery time must be defined, according to the characteristics of the application, and taking into account the implementation cost of the consistency check as well.

A second issue is the definition of the consistency check procedure to be adopted. Considering that the consistency check aims to detect the occurrence of faults affecting the correct operation of the system, the consistency check plays an important role in the achievement fault tolerance.

In the definition of the processor context, designers must identify the minimum set of information that is necessary to allow the system to be restored to an error-free state when a fault is detected. The amount of data to be saved affects the time required to perform checkpoints and also rollback. Therefore, to achieve low performance overhead during normal operation, as well as fast recovery, the minimum transfer time for those operations must be obtained, together with a low implementation cost.

As an example of task-level fault detection scheme, we can consider the approach presented in (Pignol, 2006). In the DT2 architecture (Pignol, 2006), two processors execute in parallel the same task, as in the lockstep architecture. However, fault detection is performed only when the two processors completed the task execution, by checking the results produced. As a result, the architecture is less intrusive in the task execution time, as few consistency checks are performed compared to lockstep. On the contrary, error detection latency is higher than in the lockstep, as consistency checks are performed only after task completion. Upon error detection, the two processors are reset, and the whole execution of the task is repeated.

3.2.1.3. Information and Time Redundancies

Fault detection and compensation techniques that exploit passive/active redundancy entail a significant cost overhead, as they require at least two copies of the same processor. Moreover, additional hardware implies more power consumption, as well as additional volume and weight, which can be very critical factors for some type of applications (like for example electronic equipment on board of small-size satellites).

To balance fault mitigation capabilities with implementation costs, alternative approaches to processor redundancy have been proposed which exploit information and time redundancy. The basic idea consists in adding additional instructions to those actually needed by the application software to cope with fault detection and compensation. By spending more time in doing the needed computation, and storing more information than that actually needed, several authors showed the possibility of detection and correcting the occurrence of transient errors. As previously done, the techniques can be broadly classified as:

- Instruction-level redundancy techniques, where the instructions (assembly, or C, etc.) of the application source code are duplicated, consistency checks on the results of instructions, and error handling functions are added.
- Task-level redundancy techniques, where the tasks the processor executes are duplicated, consistency checks on the results of tasks, and error handling functions are added.

3.2.1.3.1. *Instruction-Level Redundancy*

Instruction-level redundancy can be roughly described as a technique that introduces additional instructions to the source code (at the assembly level, or at higher levels) of an application to perform error detection and compensation. The added instructions replicate computations, data structures, or perform consistency checks (of stored data, or taken branches) and, if needed perform error correction operations. For the sake of clarity it is convenient to distinguish among data-oriented techniques, to detect and compensate transient errors that affect the data the program manipulates, and control-oriented techniques that deal with transient errors affecting the program execution flow.

A basic method to achieve error detection by data and code redundancy exploiting a set of transformations on the high-level source code is presented in Golubeva et al., 2006. Errors affecting both data and code can be detected by the hardened code, where each variable is duplicated, and consistency checks are added after every read operation. Other transformations focusing on errors affecting the code are devised, corresponding from one side to duplicating the code of each write operation, and from the other to adding consistency checks after the executed operations. Placing the check operation after every read operation allows reducing the effects of possible error propagations.

A major advantage of the method lies in the fact that it can be automatically applied to a high-level source code, thus freeing the programmer from the burden of performing this task, and guaranteeing the correctness and effectiveness of the resulting code (e.g., by selecting what to duplicate and where to put the checks). The method is independent from the hardware that will run the code, and it can be combined with other already existing error detection techniques.

The transformation rules affect the variables of the program. By acting on the source code they are independent on whether the variables are stored in the main memory, in a cache, or in a processor register. The fault detection is quite high, as the transformation rules do not target any specific fault model.

The transformation rules are the following:

- Rule #1: every variable x is duplicated (let x_0 and x_1 be the names of the two copies)
- Rule #2: every write operation performed on x is performed on both x_0 and x_1
- Rule #3: each time x is read, the two copies x_0 and x_1 are checked for consistency, and an error detection procedure is activated if an inconsistency is detected.

Fault propagation is blocked by performing the consistency check immediately after the read operation. Variables should also be checked when they appear in any expression used as a condition for branches or loops, in order to detect errors that corrupt the correct execution flow of the program.

Any fault affecting a variable during program execution is detected as soon as the variable becomes the source operand of an instruction, i.e., when the variable is read; this minimizes error latency, which is equal to the temporal distance between the fault occurrence and the first read operation. Errors affecting variables after their last usage remains undetected (but they are harmless, too). Figure 5 reports an example of application of the described rules: a code fragment is shown, together with its hardened version.

To understand the methods developed to cope with errors affecting the program execution flow (named *Control Flow Errors* or *CFEs*), it is important to understand first how to represent the structure of a program.

A program code can be partitioned into basic blocks (BBs), which are maximal sequences of consecutive program instructions that, in absence of faults, are always executed altogether from the first to the last. A BB does not contain any instruction changing the sequential execution, such as jump or call instructions, except for the last one. Furthermore, no instructions within the BB can be the target of a branch, jump or call instruction, except for the first one. A BB body corresponds to the BB after the removal of the last jump instruction, if this is a jump instruction. Any program P can be represented with a Control Flow Graph (CFG) composed of a set of nodes V and a set of edges B, P = {V, B}, where V = {v_1, v_2, …, v_n} and B = {$b_{i1,j1}$, $b_{i2,j2}$, …, $b_{im,jm}$}. Each node $v_i \in V$ corresponds to a BB. Each edge $b_{i,j} \in B$ corresponds to the branch from node v_i to node v_j.

Considering the CFG P = {V, B}, for each node v_i we define *suc*(v_i) as the set of BBs which are successors of v_i and *pred*(v_i) as the set of BBs which are predecessors of v_i. A BB v_j belongs to *suc*(v_i) if and only if $b_{i,j} \in B$. Similarly, v_j belongs to *pred*(v_i) if and only if $b_{j,i} \in B$.

A branch $b_{i,j}$ is *illegal* if $b_{i,j} \notin B$. If a fault causes the program to execute a branch $b_{i,k} \in B$ instead of the correct branch $b_{i,j}$, then the branch $b_{i,k}$ is *wrong*. Illegal and wrong branches represent CFEs.

All instruction-level techniques to detect CFEs are based on the idea of adding instructions to verify whether transitions among BBs are consistent with the CFG (Golubeva et al., 2006). We can consider the method named Enhanced Control flow Checking using Assertions (ECCA) as an example of this idea.

ECCA assigns a unique prime number identifier (the *Block Identifier* or *BID*) greater than 2 to each BB of a program. During program execution the global integer variable *id* is updated to store the currently traversed BID.

Each code block is modified by adding two assertions to it:

- A SET assertion at the beginning of each BB, performing two tasks: it assigns the BID of the current block to the *id* variable and then checks if the execution came from is a predecessor BB, according to the CFG. As these tasks are performing resorting to arithmetic operations, a divide-by-zero trap is executed in case of CFE.

Figure 5.a) A code fragment; b) The hardened version of the code fragment

```
…
alfa = beta*k[i]+omega*w[j];
…
```

(a)

```
…
alfa0 = beta0*k0[i0]+omega0*w0[j0];
alfa1 = beta1*k1[i1]+omega1*w1[j1];
assert( alfa0 == alfa1 );
assert( beta0 == beta1 );
assert( k0 == k1 );
assert( i0 == i1 );
assert( j0 == j1 );
assert( w0 == w1 );
…
```

(b)

Figure 6.a) A code fragment; b) The hardened version of the code fragment

```
…
alfa = beta*k[i]+omega*w[j];
…
```

(a)

```
…
SET( id, BID_current );
alfa = beta*k[i]+omega*w[j];
TEST( id, BID_current, BID_succ );
…
```

(b)

- A TEST assignment at the end of each BB, performing two tasks: it updates the *id* variable taking into account the whole set of successors according to the CFG and then checks if the current value of the *id* variable is equal to BID.

Figure 6 reports a code fragment and its hardened version based on ECCA.

When inserting data- and control-oriented fault detection and compensation techniques in a program, particular care must be placed in preserving its CFG. Indeed, if additional branches are added as a consequence of the application of fault detection and compensation techniques, for example branches hidden in the assert() code used in Figure 5b, the structure of the resulting program graph will be different from the original one, and new CFEs are introduced. As a result consistency checks must be performed in such a way that the PG of the hardened program remains the same of the unhardened one, i.e., by using arithmetic functions that produce a divide-by-zero trap in case of error detection.

The adoption of information and time redundancies entail enriching the program with a number of additional instructions, as a result program execution time is reduced. Moreover, the compiler optimization options that modify the program structure should be disabled, otherwise the PG of the executable code will different from that of the source code, and hence the fault detection and compensation techniques will be ineffective. As a result, the execution time of the program obtained after information and time redundancies application is significantly increased with respect to that of the un-hardened program. This problem can be partially solved by demanding some of the operation of the robust program to special-purpose hardware, which is in charge of accelerating the most recurring (and thus time consuming) operations that information and time redundancies perform (Bernardi et al., 2006).

An example of real-life application of instruction-level redundancy can be found in the Proton series of single board computers manufactures by SpaceMicro (www.spacemicro.com), which exploits the patented Time Triple Modular Redundancy (TTMR). According to TTMR, each instruction is executed twice, being A_1 and A_2 the two instances of instruction A, and the two results compared. In case of mismatch, a third instance A_3 of the instruction is executed, and the result to be committed to the rest of the application is obtained by majority voting among the results of A_1, A_2, and A_3.

3.2.1.3.2. *Task-Level Redundancy*

Task-level redundancy techniques combine data- and control-oriented fault detection and compensation techniques by executing twice the same task, and then comparing the obtained results. In case of single-core processor, the two instances of the same task are executed serially, while in case of dual-core processors, they can be executed concurrently.

An example of task-level redundancy can be found in the DMT architecture (Pignol, 2006). A single-core processor executes two instances of the same task, which are assigned to two different, not overlapping, and isolated memory spaces.

When the results of the two tasks are available, a hardware module compares them, and in case of mismatch resets the processor, and force a new execution of the whole process.

3.2.1.4. Fault Coverage and Overhead Analysis

When considered the fault coverage capabilities of the different techniques we can conveniently talk about fault detection, defined as the percentage of faults the method is able to detect over the faults that may hit the processor, and fault correction, as the percentage of faults the method is able to recover from over the faults that may hit the processor. In Table 1, we report a fault coverage analysis for different techniques, assuming that a processor such as the PPC750FX is considered. Faults in the memory elements not protected by parity or ECC are considered, only (i.e., faults in user registers and control registers and faults in cache memory are neglected).

When analyzing the overhead of the considered techniques we should take into account the hardware overhead (e.g., the number of additional processors, and the area of the additional hardware expressed as a fraction of the processor area), the additional memory occupation, and the additional program execution time. Table 1 compares the overhead of the different architecture normalized versus a system that does not include error detection and compensations techniques. For computing the figures we assume a complex processor like PPC750FX, whose area is notably bigger than that of additional hardware that may be needed by fault detection and compensation techniques.

As far as the fault coverage is concerned, we can see that even the original, not robust, system is able to detect some of the faults that radiations may induce. Indeed, processors normally embed some error detection features (beside parity or ECC for memory banks like caches) that can provide partial transient error detection. As an example some faults may be detected through processor traps (e.g., invalid op-code trap). All methods offer complete fault coverage, thus guaranteeing at least the capability of recognizing an improper processor behavior. Among the

Table 1. Overhead comparison

Technique	Fault detection [%]	Fault correction [%]	Area occupation [#]	Memory occupation [#]	Execution time [#]
None (original system)	<10	0	100	100	100
Passive redundancy (TMR) to perform error masking	100	100	<305	<105	<105
Active redundancy (lockstep, with checkpoint executed only in correspondence of 10% of memory accesses) to perform error detection and removal	100	98	<210	<205	<220
Active redundancy (DT2) to perform error detection and removal	100	100	<205	<205	<210
Instruction-level redundancy, with data- and control-oriented techniques to perform error detection only	100	0	100	<400	<300
Instruction-level redundancy, with data- and control-oriented techniques+hardware acceleration to perform error detection only	100	0	<110	<210	<205
Task-level redundancy (DMT) to perform error detection and removal	100	100	<110	<210	<400

methods that offer fault compensation, fault correction is at least 98%, thus guaranteeing the possibility of resuming the correct operations upon fault detection of the vast majority of faults.

As far as area occupation is considered, all methods except purely-software ones entail some additional hardware, with complexity ranging from small watchdog up to one or two instances of the original processor.

Considering memory occupation, purely-software techniques are the most demanding ones, due to the replicated instructions a data. Similarly, software techniques are the most demanding as far as execution time is considered. Other techniques like DT2 and DMT show high execution time overhead, as a slack time equal to the original program execution time (DT2) or twice of the program execution time (DMT) are reserved for error compensation through task re-execution.

3.2.1.5. Final Remarks

The detection and compensation techniques to be used should be selected on the basis of the application the system is aiming at by evaluating carefully the magnitude of the problem transient faults induce. In certain type of applications transient errors may result in the modification of one pixel, and may occur once every year; as a result, it may be deemed more convenient to avoid any mitigation technique thus saving implementation costs. In other type of application, like the rendezvous with a comet, the system may have to remain operative only few minutes, during which no error is acceptable, otherwise the entire mission is lost; as a result, the most expensive fault detection and compensation technique should be used to guarantee 100% system availability.

Moreover, there is not a single best technique that solves all the problems. Certain types of effects induced by transient errors are either too expensive, or impossible to deal with certain techniques, and therefore a mix of different techniques is likely to be the best solution. Therefore, the general architecture of a reliable processor module should entail the processor that executes the application software, possibly enriched with information and time redundancies, plus additional hardware devoted to fault detection and compensation. This module can be as simple as a watchdog timer to reset and wake-up the processor that entered into end-loop due to very critical transient faults, or as complex as the hardware needed for implementing an NMR architecture.

3.2.2. Memories

Memories are increasingly critical components from the point of view of reliability. Their criticality is due to several factors:

- The high sensitivity of some semiconductor memory technologies (e.g., SRAMs) to radiations; in particular, CMOS based memories are especially prone to radiation-induced errors (Gong, 2003). This is mainly due to the fact that design rules are typically very aggressively pushed when designing memory arrays for area optimization. Moreover, the transistors used in memory cells are often optimized for area and speed and are therefore more sensitive to noise as compared to their logic counterparts.
- The high amount of memory existing in most of today processing systems; many complex Integrated Circuits currently on the market use a high percentage (from 50% to 70%) of the die area for embedded memory cores. It is projected that this percentage will soon increase to 90% of overall chip area on more demanding SoC designs. On the other side, the amount of semiconductor memory existing in most general-purpose systems (e.g., Personal Computers or workstations) already exceeds several GByte even for low-cost

products, making the frequency of faults affecting memories extremely high.

For the above reasons, several techniques for detecting and possibly correcting errors affecting memories have been developed in the past decades, and are now commonly adopted in many commercial products, even for non safety-critical applications (Slayman, 2005). However, the choice of the solution which provides the best trade-off between fault detection/correction capabilities, area and performance cost may significantly change from case to case, depending from many factors, such as the application characteristics, the available technology, the environment the product should work in, etc.

In the following, a summary of the most common techniques used to protect semiconductor memories against transient faults is presented. We did not consider here solutions based on special memory cells suitable designed for high reliability, which are out of the scope of this chapter.

3.2.2.1. Information Redundancy Techniques

This group of techniques is mainly based on error detection and correction codes, which are used extensively in memory arrays in order to reduce the impact of transient errors, whichever their origin. The main idea behind these codes is to use check bits in addition to the data bits to create a *checkword*. The simplest form of error detection is the use of a parity bit. Each time a word is written to the memory, an extra bit is computed and stored in the memory, which represents the binary sum (or its complement) of all the bits in the word (even or odd parity). If any bit (either in the data part or in the code) is flipped, the parity bit will no longer match the data and an error is detected when the word is read. However, in the case an error is detected the technique cannot provide the information about which bit has been flipped, so the error cannot be corrected.

Several solutions exist to implement detection mechanisms based on parity; they differ on which data bits are associated to each parity bit. When the whole memory is composed of different modules in parallel, it may be convenient to associate each parity bit either to a whole word, or to the bits of the word implemented by each module, or to a proper combination of them. The major advantage coming from these schemes, which are summarized in Figure 7, is that they allow detecting errors arising when a whole memory module becomes faulty. Finally, there have been some efforts to extend the parity method to detection of multiple bits (no matter their multiplicity) by proposing 2D schemes, such as the one shown in Figure 8 (Pflanz, 2002).

The adoption of more check bits per word allows getting more powerful detection and correction codes. Single error correct/double error detection (SEC-DED) codes were first introduced by Hamming (Hamming, 1950) and then continuously improved in the following years (Chen, 1984).

The number of check bits required to achieve a given goal (e.g., detection/correction of errors with a given multiplicity) is a function of the size of the data to be protected: this means that the percent area overhead for storing the check bits decreases when the length of the protected word increases. In some cases (e.g., for caches) entire lines are protected with a very reduced number of additional check bits, when compared to the number of data bits. In other cases (e.g., small caches) lines are much smaller, and the overhead required by the codes may be unaffordable: in this case, parity is often adopted, in conjunction with more complex techniques: in some cases, the cache line is simply flushed when an error is detected, and its correct content is then uploaded from memory.

Double-bit error correction/triple-bit error detection (DEC-TED) is another powerful error correction code. Obviously, the greater error detection/correction capabilities of these codes have a

Figure 7. Different parity schemes

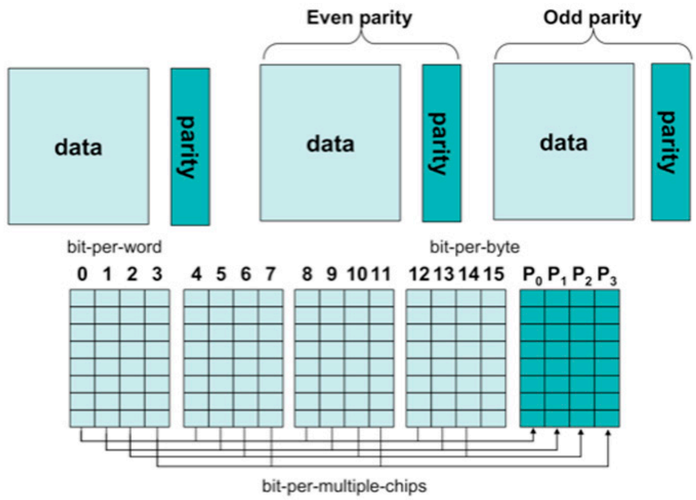
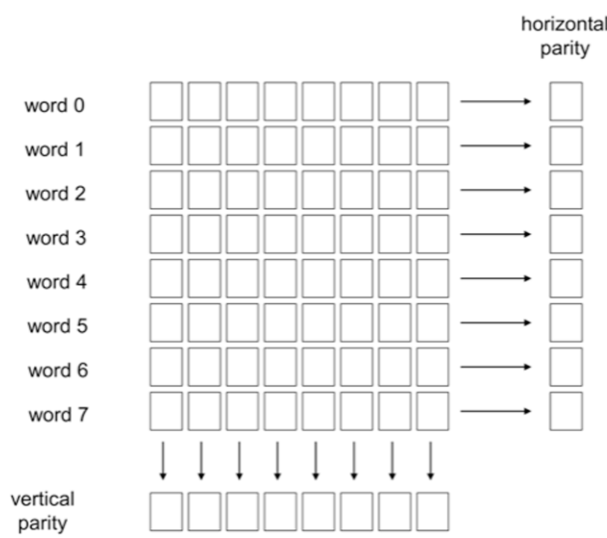


Figure 8. The cross-parity checking scheme



cost in terms of chip area, power required for the extra check bits, and latency for the computation required to encode and decode the check bits each time a word is written/read to/from memory. While sometimes adopted for space applications, DEC-TED codes are not commonly used in terrestrial workstation or server designs for several reasons. First of all, increased computational complexity leads to increased latency. Moreover, none of the word lengths, which minimize the cost for DEC-TED codes, are integral multiples of 8, thus forcing the designer to sub-optimal choices.

A further technique which is sometimes used to introduce error detection and correction capabilities to semiconductor memories is based on Reed-Solomon codes.

A Reed-Solomon code is characterized by the value of two integer parameters n and k (Blahut, 1983), where n is the number of symbols of m bits (with $n \leq 2^m - 1$) of a codeword and k is the number of symbols of the related dataword. A Reed-Solomon (n,k) code can correct up to $2E + S \leq n - k$ erroneous bits, where E is the number of erasures and S is the number of random errors. An erasure is an error whose location is known a priori (e.g., because it stems from a permanent fault). Transient faults (e.g., SEUs) can occur in unknown locations (i.e., bits) of a codeword; therefore, they can be effectively considered as random errors. Permanent faults (e.g., stuck-at 0/1 cells) can be located using either self-checking circuits, or on-line testing; therefore, they can be effectively considered as erasures.

Reed–Solomon coding is commonly used in data transmissions, as well as in mass storage systems to correct the burst errors associated with media defects. Reed–Solomon coding is a key component of the compact disc technology: it has been the first case of adoption of a strong error correction coding technique in a mass-produced consumer product, and DAT and DVD devices use similar schemes.

Reed–Solomon coding can also be used to harden semiconductor memories (in particular ROM, RAM and Flash) by adding a proper encoding/decoding circuitry on top of the memory module itself. The adoption of this technique may be convenient when the memory has to be used in particularly harsh environment (e.g., space), and has been boosted by the availability of efficient and compact Intellectual Property cores implementing highly customizable encoder/decoder functions.

Error detection latency is a major issue when designing a memory. In fact, latent errors can occur in both SRAM and DRAM due to bad writes (e.g., coming from noise in the system) or cell upsets from alpha particles or cosmic rays. These errors will remain undetected if no reads from the word occur. With a single-bit error already lying dormant, the check word is exposed to the risk of a second bit being upset, leading to an uncorrectable error in an SEC-DED design. In this case, scrubbing (Saleh, 1990) can be successfully exploited, if the application constraints allow its adoption. Scrubbing is a technique to prevent independent single-bit events from evolving into multibit events in memories: it is based on periodically checking (or scrubbing) the memory for correctable errors by traversing the entire address space. In the case of parity-protected caches, if a parity error is encountered, the data would be flushed and a re-read of correct data from main memory would occur. For SEC-DED-protected memory, a single-bit error would be detected and corrected. The probability of two statistically independent single-bit errors affecting the same word is inversely proportional to the time allowed for them to combine (i.e., the time between scrubs). Scrubbing eliminates this risk by removing the first error before the second one can arise. Clearly, scrubbing may significantly reduce the probability of failures by introducing some performance related to the time cost of the periodic scrubbing procedure.

3.2.2.2. Hardware Redundancy Techniques

The usage of triple modular redundancy even for memories is not uncommon for mission-critical avionic and space applications. On the other side, this technique is seldom used within commercial servers and workstations (and more in general for ground-level applications) because of the obvious power and area penalties. In almost all cases, judicious use of a ECC-protected cache and memory, combined with a good analysis of the Soft Error Rate (SER) of the different components, a careful design of logic elements, and a correct process-technology selection should be sufficient for commercial designs.

3.2.3. Random Logic

As described in the previous subsections, plenty transients effects may corrupt the circuits functionalities. Particular SEU and SET tolerance techniques must be taken in account in the design of fault tolerant integrated circuits.

3.2.3.1. Hardware Redundancy

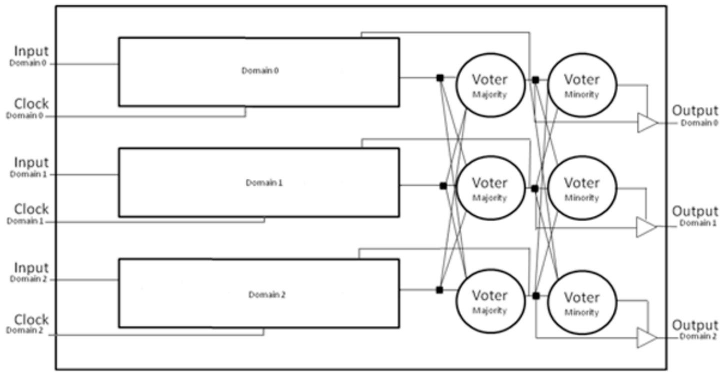
Hardware redundancy solutions can be adopted to protect the logic against transient faults. A commonly adopted solution is the Triple Modular Redundancy (TMR) which is generally used to provide design hardening in the presence of SEUs. All the circuit's flip-flops are replicated three times and a voter structure is used to the majority of the outputs, accordingly to the TMR scheme. The classical TMR scheme has been proved to be effective in protecting sequential logic against SEUs, nevertheless it brings high overhead on area. TMR techniques are based on the idea that a circuit can be hardened against SEUs or SETs by designing three copies of the same circuit and building a majority voter on the output of the replicated circuit. Implementing triple redundant circuits against SEU effects is generally limited in protecting only the memory elements. Conversely, in case SETs effects are addressed, full module redundancy is required because all the components such as memory elements, interconnections and combinational gates are sensible to SETs. This means that three copies of the user's design have to be implemented to harden the circuit against SEUs. All the logic resources are triplicated in three different domains according to the traditional TMR approach and voter elements are included at the domain's outputs. When any of the design domains fails, the other domains continue to work properly, and thanks to the voter, the correct logic passes to the circuit output. The TMR traditional architecture presents two major drawbacks:

1. The design is still vulnerable to SEUs in the voting circuitry. Nevertheless radiation-hardened voting circuits are available to protect against the permanent damaging of the voter induced by Single-Event Latch-Up (SEL), the CMOS devices adopting physical layout geometries smaller than 0.25 μm are still susceptible to transient effects.
2. There is no protection against the accumulation of SEUs in Flip-Flops (FFs). After an SEU happens, in the traditional TMR scheme, while the state machine is corrected through scrubbing, the state remains un-reset for the synchronization.

In order to protect the voters and input signals, all inputs, outputs and voters are triplicated, which avoid the presence of single points of failure. Besides, in order to ensure constant synchronization between the redundant state registers, majority voters with feedback paths are inserted. The resulting feedback logic of each state machine becomes a function of the current state voted from all the three state registers. The presence of the minority voters allow to protect the circuit outputs from SEUs affecting their resources, that represent a critical point, since they are directly connected to the device pins, as illustrated in Figure 9. The full TMR solution could not be the most harden solutions if critical harsh environments are considered. In details, if a higher level of protection against SEUs is necessary, several voter levels can be introduces into the circuit's random logic in order to increase the protection against transient effects. Nevertheless these solutions introduce a high routing congestion due to the interconnections coming from three domains to a single logic element. This has a severe impact on the delay of a design and the performance degradation may vary from 1.5 and 2.5 times with respect to the original un-hardened circuits.

A different solution, that reduces the dramatic overhead of TMR, is the Double Modular Redundancy (DMR). The main characteristic of DMR are the duplication of the storage cells and the correction elements used to output and keep the correct value of the redundant storage cells. The impact of the DMR structure on the area and on the delay is lower than TMR, while the tolerance

Figure 9. TMR architecture based on majority and minority voters



to SEU effect has similar level, in particular when sequential logic circuits are considered.

The principal DMR structure consists in doubling the storage cells of each bit and connecting the replicated outputs to the correction element. In case one of the redundant cells is changed by an SEU, the output of the correction element will maintain the right value. Then, at the next clock's edge, both the doubled storage cells will latch new data removing the possible errors. The DMR hardening solution allows saving one storage cell for each bit, therefore it is quite obvious that the area and the latency of the correction element is much less than a TMR-based voter. As a drawback, the DMR requires a physical design of the correction element, since the correction element used in DMR solution is a modified inverted latch, since the standard (symmetric) element are not able to decrease the SEU probability.

3.2.3.2. Temporal Redundancy

The traditional TMR structure that triplicates only the flip-flops can protect sequential logic against SEUs, however it cannot tolerate transient effects such as SETs in combinational logic. While a possible solution is to triplicate also the combinational paths, thus implementing a full-TMR approach, it does not protect completely the implemented circuits, since a SET glitch may pass through logic states and if multiple glitches arrive at the input of registers exactly at the clock edge, the redundant storage cells would latch incorrect values or go to undefined and metastable states. Therefore particular techniques have been developed to face SET effects. There are two common methods used for SET tolerant circuits, both are based on temporal concepts, pure temporal redundancy approach and SET filter.

The temporal redundancy (Mavis, 2002) is based on the TMR structure but it uses three shifted clocks to sample data at different time intervals. The approach is valid in case the SET glitches are shorter than the time interval, since at most one latch would sample incorrect value. A voter is used to forward the correct value to the output latch controlled by another clock domain.

The SET filter techniques are based on delaying the signals between sequential elements (Mongkolkachit, 2003). The idea behind SET filtering is to double the logical paths defining two signals: the original signal and the delayed signal, which are connected to the same correction element (which generally consists of a guard gate). Therefore SET pulses having a width shorter than the delay between the two signals would not be able to change the output of the correction element. In relation to the SET filtering, a particular kind of flip-flop has been developed: the DF-DICE cell structure is proposed to improve the aforementioned SET filter technique (Naseer, 2005).

Innovative techniques based on the broadening and filtering capabilities of logic gates are currently under evaluation. The idea behind such techniques is based on preventing the effect of propagation of SETs in combinatorial circuits. Two approaches are nowadays proposed: the first is oriented in the modification of the gate design to present balanced high to low and low to high propagation delays (Kastensmidt Lima, 2008), the second is oriented to the modification of the routing electrical characteristics by modifying the placement position of each logic cell within the integrated circuits. Both the approaches presented encouraging results.

4. CONCLUSION

The increasing usage of electronic systems for safety critical applications, and the growing sensitivity of semiconductor technologies to transient faults pushed towards the development of different techniques for developing devices and systems able to tolerate such a kind of faults. This chapter provides an overview of the main techniques that have been proposed and/or adopted for processors, memories and logic components. It is up to the designer, based on the specific constraints of its application and design flow, to combine and implement them in the proper way.

REFERENCES

Abate, F., Sterpone, L., & Violante, M. (2008). A New Mitigation Approach For Soft Errors In Embedded Processors. *IEEE Transactions on Nuclear Science*, *55*(4), 2063–2069. doi:10.1109/TNS.2008.2000839

Barak, J., Levinson, J., Victoria, M., & Hajdas, W. (1996). Direct process in the energy deposition of protons in silicon. *IEEE Transactions on Nuclear Science*, *43*(6), 2820–2826. doi:10.1109/23.556872

Barth, J. L., Dyer, C. S., & Stassinopoulos, E. G. (2004). Space, Atmospheric, and Terrestrial Radiation Environments. *IEEE Transactions on Nuclear Science*, *50*(3), 466–482. doi:10.1109/TNS.2003.813131

Baumann, R. C. (2005). Single events effects in advanced CMOS Technology. In *IEEE Nuclear and Space Radiation Effects Conference Short Course Textbook*.

Bernardi, P., Veiras Bolzani, L. M., Rebaudengo, M., Sonza Reorda, M., Vargas, F. L., & Violante, M. (2006). A New Hybrid Fault Detection Technique for Systems-on-a-Chip. *IEEE Transactions on Computers*, *55*(2), 185–198. doi:10.1109/TC.2006.15

Blahut, R. E. (1983). *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley Publishing Company.

Chen, C. L., & Hsiao, M. Y. (1984). Error-correcting codes for semiconductor memory applications: A state-of-the-art review. *IBM Journal of Research and Development*, *28*(2), 124–134. doi:10.1147/rd.282.0124

Dodd, P. E., & Massengill, L. W. (2004). Basic Mechanism and Modeling of Single-Event Upset in Digital Microelectronics. *IEEE Transactions on Nuclear Science*, *50*(3), 583–602. doi:10.1109/TNS.2003.813129

Ferlet-Cavrois, V., Paillet, P., McMorrow, D., Fel, N., Baggio, J., & Girard, S. (2007). New insights into single event transient propagation in chains of inverters-evidence for propagation-induced pulse broadening. *IEEE Transactions on Nuclear Science*, *54*(6), 2338–2346. doi:10.1109/TNS.2007.910202

Golubeva, O., Rebaudengo, M., Sonza Reorda, M., & Violante, M. (2006). *Software-Implemented Hardware Fault Tolerance*. Berlin: Springer.

Gong, M. K. (2003). Modeling Alpha-Particle-Induced Accelerated Soft Error Rate in Semiconductor Memory. *IEEE Transactions on Electron Devices*, *50*(7), 1652–1657. doi:10.1109/TED.2003.813905

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal*, *29*(2), 147–160.

Kastensmidt Lima, F., Wirth, G., & Ribeiro, I. (2008). Single Event Transients in Logic Circuits-Load and Propagation Induced Pulse Broadening. *IEEE Transactions on Nuclear Science*, *55*(6), 2928–2935. doi:10.1109/TNS.2008.2006265

Mavis, D. G., & Eaton, P. H. (2002). Soft error rate mitigation techniques for modern microcircuits. *Reliability Physics Symposium Proceedings*, 40th Annual, (pp. 216-225).

May, T. C., & Woods, M. H. (1979). Alpha-particle-induced soft errors in dynamic memories. *IEEE Transactions on Electron Devices*, *26*(2), 2–9. doi:10.1109/T-ED.1979.19370

Mongkolkachit, P., & Bhuva, B. (2003). Design technique for mitigation of alpha-particle-induced single-event-transients in combinational logic. *IEEE Transactions on Device and Materials Reliability*, *3*(3), 89–92. doi:10.1109/TDMR.2003.816568

Naseer, R., & Draper, J. (2005). The DF-DICE storage element for immunity to soft errors. In *proceedings IEEE Midwest Symposium on Circuits and Systems*.

Pflanz, M., Walther, K., Galke, C., & Vierhaus, H. T. (2002). On-Line Error Detection and Correction in Storage Elements with Cross-Parity Check. In *Proceedings of the Eighth IEEE International On-Line Testing Workshop,* (pp. 69-73).

Pignol, M. (2006). DMT and DT2: Two Fault-Tolerant Architectures developed by CNES for COTS-based Spacecraft Supercomputers. In *Proceedings of the 12th IEEE International On-Line Testing Symposium*, (pp. 203-212).

Pradhan, D. K. (1996). *Fault-Tolerant Computer System Design*. New York: Prentice Hall.

Rajeevakumar, T. V., Lu, N., Henkels, W., Hwang, W., & Franch, R. (1988). A new failure mode of radiation-induced soft errors in dynamic memories. *IEEE Electron Device Letters*, *9*(12), 644–646. doi:10.1109/55.20423

Saleh, A. M., Serrano, J. J., & Patel, J. H. (1990). Reliability of scrubbing recovery-techniques for memory systems. *IEEE Transactions on Reliability*, *39*(1), 114–122. doi:10.1109/24.52622

Slayman, C. W. (2005). Cache and Memory Error Detection, Correction, and Reduction Techniques for Terrestrial Servers and Workstations. *IEEE Transactions on Device and Materials Reliability*, *5*(3), 397–404. doi:10.1109/TDMR.2005.856487

Sterpone, L., Battezzati, N., Ferlet-Cavrois, V. (2009) Analysis of SET Propagation in Flash-based FPGAs by means of Electrical Pulse Injection. *IEEE Radiaction Effects on Component and Systems*, 2009.

Taber, A. H., & Normand, E. (1995). *Investigations and characterization of SEU effects and hardening strategies in avionics*. Alexandria, VA: Defense Nuclear Agency.

Troutman, R. R. (1983). Epitaxial layer enhancement of n-well guard rings for CMOS circuits. *IEEE Electron Device Letters*, *4*(12), 438–440. doi:10.1109/EDL.1983.25794

This work was previously published in Design and Test Technology for Dependable Systems-on-Chip, edited by Raimund Ubar, Jaan Raik and Heinrich Theodor Vierhaus, pp. 132-154, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Chapter 97
Augmented Reality for Collaborative Assembly Design in Manufacturing Sector

Rui (Irene) Chen
The University of Sydney, Australia

Xiangyu Wang
The University of Sydney, Australia

Lei Hou
The University of Sydney, Australia

ABSTRACT

Some speculations regarding the current issues existing in current assembly work have been brought out in this chapter. The theoretical basis behind the idea that Augmented Reality systems could provide cognitive support and augmentation is established. Hence, this chapter analyses the foundations for assembly feasibility evaluation and discusses possible innovative ways to provide an efficient and robust solution for these problems of realism and efficiency in design and assembly processes. The proposed platform considers the multiple dependencies in different manufacturing sectors that allow the work to be conducted in a simultaneous way, rather than sequential order.

INTRODUCTION

For a typical machine, one may look inside and find that it consists of accessories, modules and parts which have to be combined and integrated to fulfill its function. As the last step of mechanical manufacture to satisfy the technical requirements, assembly also includes the relative works such as regulation of assembly accuracy, inspection of tolerance, test of integration, painting and package other than the pure combination of components. During the process, correctly guaranteeing the assembly precision of machine and components is one of the critical criteria determines the product quality. Before shaping the finished product, the traditional assembly of mechanical components is mainly on the basis of parameters deriving from assembly drawings or some business software

DOI: 10.4018/978-1-4666-1945-6.ch097

such as CAD/CAM. Although being guaranteed a realization of strict assembly sequence and considerable assembly precision, such methods guided by assembly drawings or software also manifest their defects. Because of the lack of combination with the real environment, they can't provide a better understanding of different constrains and complicated plant environments in the real assembly circumstances while the after evaluation of assembly workload such as assembly cost, time consuming and hardship could not be easily commanded either. Moreover, concerning a large amount of components to be assembled, an obstacle to understand the assembly drawing between the designers and the assemblers has been introduced. Again, the repetitive assembly operation would also create a incorrect assembly because of the physical tiredness and the mental negligence. Accordingly, such a serious occlusion of information exchange greatly limits the development of the assembly level. To solve this problem, a concept of collaborative assembly design by using AR technology has been arisen, aiming at completing the interconnection of every character in the before-product period. This technology will allow designers, manufacturers and assemblers to collaborate in a distributed way.

This paper mainly focuses on a novel technology to evaluate mechanical assembly sequences, say, the Augmented Reality technology. Defined as the combination of the real and virtual scene, such a technology has been widely used in areas such as product maintenance (Feiner & MacIntyre & Seligmann, 1993), manufacturing (Curtis & Mizell & Gruenbaum & Janin, 1998; Sims, 1994), training (Back, 2001), battlefield (Urban, 1995; Metzger, 1993), medicine (Mellor, 1995; Uenohara & Kanade, 1995), and so on (Pope, 1994). Comparing with Virtual Reality (VR) which aims at totally mimicking the real world into the synthetic virtual perception according to the help of computer and separates the real and virtual environment, AR maintains a sense of presence in the real world and balances the perception of real and virtual world. With the help of a see-through head-mounted device, a data glove and a tracking system, the virtual components can be placed in the real one which can be seen from the head-mounted device. Wearing eyeglasses which are also called the HMD, an operator manipulates the virtual components directly and superimposes them on the real ones with the whole process can be seen from the HMD. Besides, during the operation process of integrating the virtual objects, the operator will detect the real interferences coming from the real objects and the ambient environment that would delay the assembly schedule or need to be further modified. What is more, because of the feedback of other "non-situated" elements like recorded voice, animation, replayed video, short tips, arrows, it could simultaneously guide the operators through the whole assembly operation, release their tension and even notify an error assembly (Raghavan & Molineros & Sharma, 1999). This way, the reality being perceived is augmented. In AR, two noticeable key techniques have to be mentioned. They are the generation of assembly sequence and the superimposition of virtual objects.

The planner, a key component in the system for creating the assembly sequence and reducing the impractical or impossible subsequences while manipulating the real and virtual prototype components in an assembly environment, allows an operator to cooperate with using the assembly graph which is also called the liaison graph (Molineros, 2002). The generation and the decrease of assembly sequence are coming from the mutual behavior between the operator and the planner. With the help of the assembly graph (liaise graph) which serves as the underlying control structure, the planner utilizes precedence constrains along with the liaison graph to input all the possible assembly sequences by planner, say, a state from FFFFF to TTTTT, and then generates all the possible sequences. Following that, a mutual intercommunication between the planner and the operator is used to reduce the impractical and impossible assembly sequences. When generating the assembly sequence, the planner initiatively provides information to the augmentation system on what to display using the liaison graph which detects the interference, the inappropriate sequences like the unmatchable inserts, connections and fixing, thus largely reducing the impractical and impossible assembly sequences. In the opposite way, according to visual sensing and augmentation, the operator provides the information perceived to the planner about what parts are being manipulated or what state has been reached in the assembly representation, simultaneously a liaison has been established when graphics and augmentation cues are correctly positioned and updated. The planner could delete the finished assembly notes in the liaison graph and reduce some assembly sequences. Again, the operator provides the information to the planner on the basis of how difficult the completed assembly sequence is and then this information is used to cumulatively compute the cost of the sequence under evaluation and to eliminate other sequences (Wilson, 1995). Another technique used for augmentation is the marker capture technique where the three algorithms were used, the thresholding algorithm, the region growing algorithm and Minimum Spanning Trees (MST) algorithm (Raghavan & Molineros & Sharma, 1999; Bolt, 1980). Considering computer vision techniques offer the potential to provide rich feedback, but suffer from problems like computational complexity, calibration, occlusion caused by lighting conditions, noise, etc. Accordingly, the assembly uses markers to simplify the computer vision algorithms and makes real-time implementation feasible.

BACKGROUND

Some assembly systems based on augmented reality have been produced by the researchers. The two most important AR systems were introduced by Molineros (2002) and by Azuma (1993), they developed AR system called the AREAS and AUDIT, in which they outlined in detail the components of their systems, the principle of each operational usages such as HMD (see-through eyeglass), Camera, planner, tracker, as well as image mixing. ESKUA which is based on tangible user interfaces and carried out by Zhang, Y, Finger, S and Behrens S, is another operational platform to interact with CAD software and replace the old CAD/CAM platform concerning the limitation of environment. The camera tracking technique rested with setting a series of markers was deeply investigated by Neumann, Cho (1996) and Liverani, Amati, Caligiana (2004) to supplement the object-tracking blank that the traditional computer graphics can't realize. To test the efficiency of the AR system, especially in the assembly in manufacturing industry, researchers have carried out a great deal of work in the last few years with the help of algorithms of automatically generating the assembly sequence (Wiedenmaier & Oehme & Schmidt & Luczak, 2001; Shin & Dunston, 2008). Despite the hardships and barriers in constraints management, especially in large batch assembly, great interest in other industries is still exhibited (Pellom & Kadri, 2003). More in general, there are some defects that should be noted. At first, when assembling in the real situation where only a few accessories should be assembled, it is better to obey the instructions in assembly drawings since their simplicity, time saving and low cost. Whereas the assembly drawings become very complicated or even can't be understood by the workers with obvious ease as soon as the machine includes a large amount of accessories. Under such circumstances, reading an assembly drawing for a long time could easily create a visual and physical tiredness which lowers the assembly efficiency in the issue. Accordingly, assembly using AR technique should be employed. But in the current AR system, we still can't utilize the computer to automate the sequence of assembly in view of the conditionality of colossal algorithm and computer Graphics (Tuceryan & Greer &

Whitaker & Breen & Crampton & Rose & Ahlers, 1995). Moreover, during the assembly, the mutual accuracy of position precision (straightness, parallelism and coaxiality), the mutual kinematic accuracy (driving precision and rotary accuracy) and the machining accuracy (matching quality and contact quality) must be guaranteed and if not, then rejects would be made in case the assembly accuracy can't be guaranteed.

ISSUES

Problem 1

To guarantee the possibility of high assembly accuracy, the proper assembly approaches should be employed. This way, it is significant to have the expert assemblers experienced in programming assembly path and solving assembly problem. To ensure accuracy in a multi-components assembly where includes lots of composed loops, an experienced expert assembler should design the ideal sequence of assembly by choosing the appropriate assembly approach among exchange, selection, modification and regulation approaches. Actually it often takes months or even years for a novice assembler to develop expert knowledge for assembling processes that have high complexity.

Problem 2

When assembling in the real situation where only a few accessories wait to be assembled, it is better to follow the instructions in assembly drawings due to their simplicity, time saving and low cost. Whereas the assembly drawings become very complicated or even can't be understood with obvious ease when the machine to be assembled includes a large amount of components. Under such circumstance, reading an assembly drawing for a long time could easily create a visual and physical tiredness which lowers the assembly efficiency in the issue. Accordingly, assembly using AR technique should be employed. But in the current AR system, we still can't utilize the computer to automate the assembly sequence in view of the limitation of computer graphics and colossal algorithm. Moreover, during the assembly, the mutual position accuracy (straightness, parallelism and coaxiality), the mutual kinematic accuracy (driving precision and rotary accuracy) and the machining accuracy (matching quality and contact quality) must be guaranteed and in case the assembly accuracy can't be guaranteed, then rejects would be made.

SOLUTIONS AND RECOMMENDATIONS

The Virtual Reality concept has been implemented in various business and industrial sectors. It offers a novel approach to human-machine interactions where information is displayed in the field of vision of human operators. This helps to visualize innovative ideas and user-activated visual programming processes. This concept also turns enterprises into a modern digital factory, particularly for the manufacturing sectors where users are often geographically distributed in different locations and collaboration is of paramount importance. Efficient sharing of product data is a crucial factor for the next generation of manufacturing.

In addition good collaboration is a key value for society. Working together in real time either together in one room or remotely is important for a flexible and fast planning process. There are a small number of existing technologies and systems which provide collaborative frameworks (Chryssolouris, et al., 2007; Constantinescu, et al., 2006) not only is collaboration between people important, but also collaboration between available systems is of significance. Important features such as common data exchange formats and a common look and feel for the given user interfaces must be considered in order to offer an easy-to-use technology that hides underlying technological diversity and complexity from users.

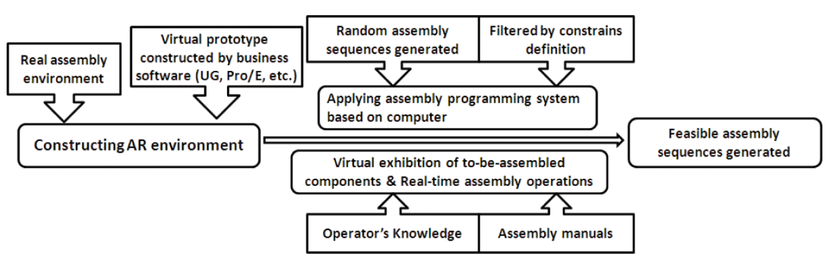
Concerning hardware ergonomics within the virtual technologies domain, a variety of devices are available which differ significantly in properties such as functionality, dimensions, weight and costs. Depending on the task and the needs of the user, the necessary hardware components have to be chosen accordingly. An example system for collaborative planning which is integrated in the Digital Factory for Human-Oriented Production System (DiFac) framework such as the IPA Collaborative Planning Table (Constantinescu & et al, 2006). It enables the collaboration of multiple involved persons, the so called "planners", for achieving the layout planning tasks. The assemblers, designers and engineers can operate with virtual objects and communicate easily and in a direct way. To collaborate with other remote users, a Multiuser Server can be employed. Files, text chat or interaction events for moving objects can be shared.

For the base components it is important to ensure the quality and robustness in order to have reliable devices for the interaction tasks. Therefore, it is envisaged that Augmented Reality (AR) systems are the most suitable for application assembly and design in manufacturing because AR ensures an efficient design, manufacture, and collaboration, as well as robust manufacturing techniques. In addition, the seamless interaction between the real and virtual world is of great importance for an easy-to-use and intuitive user interface. Again, tracking accuracy and precise registration of digital data are key issues. This chapter discusses the possibility to build an interaction platform which integrates different input and output devices like a Holobench System, DataGlove and Tracking System. Given such an interaction platform, the user is provided with a user-friendly interface for performing different tasks throughout the Digital Factory, such as designing a workflow, through direct manipulation by virtual human metaphors.

One of most difficult tasks for engineers is the assembly sequence planning for complex products (Bolt, 1980). Many automatic systems have been used to automate the sequences of the planning process, however, the problem is that it is very hard to formalize the assembly planners' knowledge. Automatic sequence planning systems are able to generate a set of feasible sequences based on identified constraints. Unfortunately, certain constrains are difficult to identify if a good and realistic feel of the assembly process is not in place. An Augmented Reality interface mixing the real and virtual prototypes provides a better perceptive environment for engineers to experience the realistic feeling of assembly operation and identify the assembly constraints. The AR assembly environment can be integrated with an automatic assembly planning system. As an example, the methodology developed by Pang et al. (2006) for assembly sequence optimization is shown in (Figure 1).

Different engineers and designers can view and verify feasible sequences in AR environments to identify new constraints and decide whether there is a need to change the optimization criteria

Figure 1. Sequence planning in AR assembly system (Adapted from (Pang, et al., 2006)).



(e.g., minimal cost, minimal number of orientations, etc.). Next, engineers can go back to the Automatic Assembly Planning System to re-plan the sequences no matter where they are located, so it is especially useful during a collaborative work.

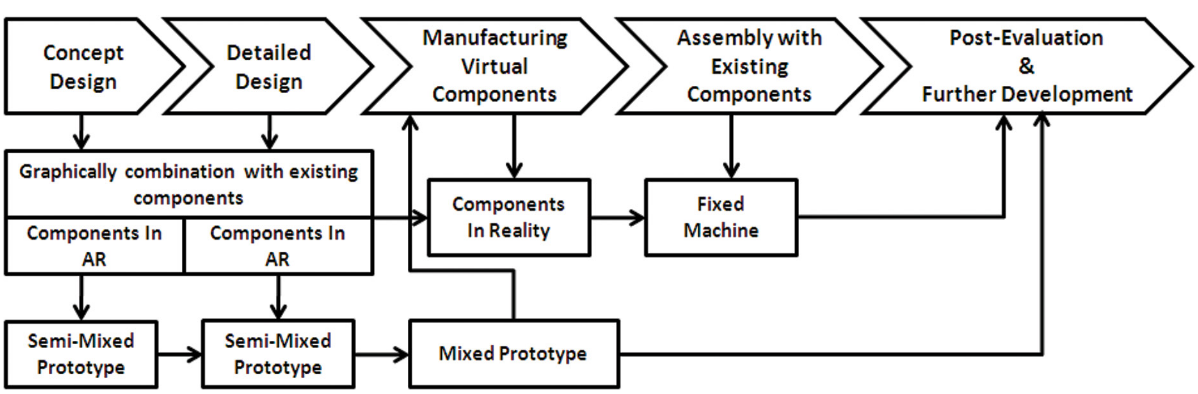
(Figure 2) describes the sequence for using the AR paradigm, starting from the concept design phase, followed by the detailed design phase, further design refinements and finally leading to building and assembly.

Concept Design phase: Business process re-engineering is increasingly becoming the focal point of attention in today's efforts to overcome problems and deficits in the automotive and aerospace industries (e.g., integration in international markets, product complexity, increasing number of product variants, reduction in product development time and cost). Deploying Augmented Reality systems often involves several interactive paradigms for assembly tasks (See Figure 2) in which engineers and designers can experiment naturally with the prototype. The figure shows a paradigm about how a single assembly can be done with the support of Augmented Reality systems. For example, the assembly process below illustrates how a designer would refer to an AR model (e.g. a design for an aircraft) to demonstrate the practicality of several designs and choose the most suitable option corresponding to existing components. An iterative process is applied in which increasing levels of detail and accuracy accumulates over the product design cycle. Initially, in the early concept design phases the AR model will be simplistic and lacking in detail (a rough draft of the desired end product). Gradually more and more component information is added and produces a more detailed design.

Detailed Design phase*:* Engineers can analyse the AR model and provide feedback. Here the AR model gives designers the freedom to quickly and easily make changes to the specifications of their model.

Modifying a real world prototype, such as an aircraft, following the traditional methodology of creating real world prototypes for each design introduces significant inefficiencies in the design and prototyping process. It can be prohibitively expensive due to costs of materials, man hours spent in and assembly of prototypes, and other costs associated with the increased timeframe required due to a slower design and prototyping process (e.g. loss of market share due to lack of competitive advantage).

Figure 2. A single interactive paradigm for an assembly task



Manufacturing Virtual Components phase: After the previous two phases the design is now starting to become relatively mature. Engineers will consult with the manufacturers of the various components (ie. component suppliers) and determine if further modifications to the AR model are required. Consultation between designers, engineers and with the component suppliers will take place in a more rapid timeframe. This promotes a more efficient form of collaboration between the interested parties.

The design cycle will move in a repetitive process of design, consultation, further design improvements, and so on. The design cycle is iterative and continually builds and improves on previous design efforts using feedback provided by engineers, component suppliers, and any other interested parties. There will be a negotiation between parties whose fields of expertise are within diverse disciplines, the AR design paradigm will facilitate better communication between these parties.

Assembly with Existing Components phase: The various parts are purchased from component suppliers and assembled by assemblers (e.g. engineers or factory workers) in order to produce the real world product based on the AR model. The product is ready to be used without need for the production of a real world prototype.

During this phase it is expected that the engineers or designers might wish to provide further feedback in order to improve the AR model so that the assemblers could follow the AR model easily rather than the schematics diagram and make the corresponding change to the real components. The virtual phototype (AR model) will help the assemblers to produce the real world product in a more efficient way, and will provide a visual representation that is more intuitive for them to follow.

Post-Evaluation & Further Development phase: After production of the prototype further testing and analysis will be carried out. This may result in further modification to the AR model. Due to the flexibility of the AR paradigm modifications to the AR model can be made more quickly.

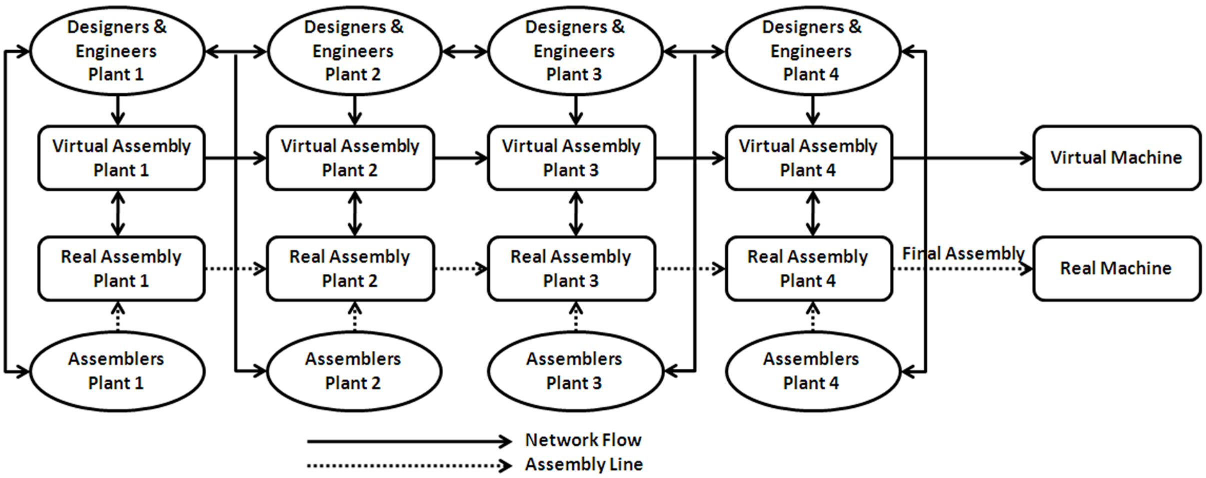
Fewer changes to the design will be required at this stage, and it will be quicker to finish the assembly tasks due to improved collaboration and feedback in earlier phases.

Practically all automotive companies separate different parts to various manufacturing factors. For example, if the object consists of four plants to be produced, then each manufacturing sector is in charge of one plant, but the important part is that the each component is required to be matched before any other steps to take further. The figure demonstrates a complicated process which using Augmented Reality concepts into a network with various departments for the components. Each sector has specialised workers in the particular domain knowledge with certain training. In the beginning the designer and engineer produce the concepts design, then with the aids of Augmented Reality systems, the virtual assembly could be formed. Meanwhile, the assemblers could see how the components should be assembled to match with some existing real components.

As an example, the assemblers of an automotive chassis in plant 2 are building the chassis based on the phototype. This plant is dependent on the production of parts from other manufacturing sectors, such as the supplier for the engine block. The manufacturer of the chassis can start to build with confidence without having to wait to receive a shipment of engine blocks from plant 1.

There are multiple dependencies between these plants, but with the network flow, each plant can monitor the other's plants' progress so there is no

Figure 3. A platform for collaborative work in assembly design simultaneously



need to wait until the previous plant to complete their assembly task, the work can be conducted simultaneously. In addition, the efficiency could be accumulated across the assembly line between all the manufacturing plants involving in the production.

BENEFITS

Reducing Mental Workload and Cognitive Load

There is a cognitive load present in the assembly process. In the assembly task, more and more assemblers get trained with different methods. Augmented Reality systems provide a more efficient way to train the assemblers, the cognitive workload will be reduced as they no longer need to exert cognitive effort when trying to visualise the component based on a traditional schematic. The AR phototype will help the assemblers by providing a more natural and intuitive phototype, requiring less effort to interpret, and helping to guide them in the assembly process.

Cognitive load theory addresses how to use elementary tools for training assembly workers. For example, users prefer a graphical user interface (GUI) that uses illuminated picture based icons in order to remember where specific programs and functionality are located in the user interface. Compare this to text based icons which are less intuitive and will increase the cognitive load. Providing visual cues decreases cognitive load even if a user knows where items are located. AR provides a superior visual stimulus to the end user. Picture based icons can be used instead of plain text icons as this provides a simpler more intuitive user interface with less cognitive load.

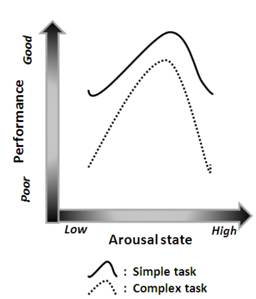
In performing complex manufacturing work such as assembly tasks, the assembler has to reach a certain level of performance while not exceeding the mental capacity he is willing to spend. A well-designed human-machine-interface like Augmented Reality might be helpful to the worker to achieve these goals. Some studies have reported that the AR environment was less mentally demanding (Tang, et al., 2003). A well-proven method of designing such systems is to arouse the workers, thus providing the visual cue is significant to the assembly task. The arousal level can be thought of as how much capacity you have available to work with. One finding with respect to arousal is the Yerkes-Dodson law (1908) which predicts an inverted U-shaped function between arousal and performance (See Figure 4). There is a certain relationship between the arousal level and the quality of performance. If the task is too simple or too complex, the level of arousal will go too low and too high. Only the optimal level of arousal, which is lower for more difficult or intellectually (cognitive) tasks, and higher for tasks requiring endurance and persistence, can improve the quality of performance. A visual stimulus is one of the significant senses for arousal and motivation of workers. In assembly tasks, workers often need to form the image of assembly sequences according to the drawings, this causes some mental workload for the workers. Although some cases may only have few steps to go through, but more often the cases are much complicated. Therefore, the Augmented Reality could allow the workers to visualize the parts rather than manipulating the objects in their mind. It has been proven that if the workers do not have to mentally transform objects and keep a model of the relationship of the assembly objects to its location in working memory, they would then experience less mental workload (Tang, et al., 2003). When the assembler is being trained to read the assembly drawings, there is a large quantity of information being presented in the drawings, it could easily overload the assembler's peak arousal level (producing stress) and this results in a situation where does not want to arouse them anymore. With the assistance of Augmented Reality systems, assemblers will not be fatigued easily since the visual aids could break down the complexity and maintain the arousal at an even level.

Reduced Design Time and Errors

Physical phototyping is traditionally the main method for assembly design. With physical phototypes, engineers, designers and assemblers can easily obtain direct feedback from their senses (visual, tactile, and force, etc) in order to decide if the parts have been designed properly. During the assembly operations, it is easier to identify any unexpected drawbacks to improve the design, but that usually is very time consuming and incurs a very high cost in past years, additionally, once being made, it is difficult, time consuming and expensive to modify the initial prototype. Hence, Augmented Reality can be potentially considered as a rapid prototyping and low cost system since the technology could generate the 3D graphical models which are used to manipulate and design products using existing (or novel) parts. Furthermore, usually the assembly designer tried to avoid the errors as much as possible. However, the design process is becoming increasingly flexible if the design process can be adopted with Augmented Reality systems since the design is changeable with changeable prototype, the 3D image will change according the manipulation from the designers so it is more efficient.

Furthermore, there seem to be two different understandings of what exactly virtual prototyping is: the "computer graphics" and the "manufacturing" point of view (Sims, 1994). During the verification between different disciplines such as designers, engineers, manufacturers and component suppliers, it is not avoidable to have some errors, some might be independent errors and some might be dependent ones. Previous study for assembly tasks have been proven that the participants who used the AR systems can produce far less dependent errors (Tang, et al., 2003). Since the visualizations provided by AR systems could help augment the assemblers' cognition and also be a more useful

Figure 4. The Yerkes-Dodson law



instructional tool compared to the paper drawings. In the past, proper design of instructional visualizations can reduce the time taken and errors made in performing a task (Molineros, 2002). To understand how to design effective visualizations, it is significant that Augmented Reality can give instant feedback, so the assembler could immediately decide if the next step should proceed before the errors to be produced.

FUTURE RESEARCH DIRECTIONS

In our future industrial case it can be aimed to enhance assembly work for sequence planning. In augmenting assembly work, the assembly worker is guided by virtual objects and visual assembly instructions.

As virtual reality has become more widespread, the demand for system developing has increased dramatically. Establishment of an AR assembly environment is an extremely complex and needs to consider many interdisciplinary issues from different domain knowledge such as design, engineering, AR technologies, assembly and manufacturing. Although there are some works which have been done in the AR assembly domain, the research in collaborative assembly design is still at the infant stage. Therefore, there are essential needs to investigate the realm of assembly process from the various disciplines in order to increase the work efficiency.

One key value for a comprehensive assembly evaluation is the physical interaction between real and virtual objects in an AR assembly environment, however there are still lots of limitations in this domain (Aliaga, 1994; Breen, et al., 19956; Vallino, 1998). For example, physical interaction between real and virtual objects in Augmented Reality should include collision detection and reaction which is significant for a comprehensive, realistic and feasible assembly and design process. If these problems cannot be solved, the AR assembly system would be hard to use in any complicated evaluation. On the other hand, at the current state of existing AR technology, the most realistic feedback can only be obtained from the direction of real object affecting the virtual components, but not vise-versa. Hence, the studies of the other way from a virtual object to the real object should be carried out as well and it has significant meaning for the future AR application in the collaboration.

CONCLUSION

This chapter discusses the current issues of assembly work and explain the reason that why the AR system could be the option to solve these problems with the potential for benefits. The proposed mixed prototyping concept from AR environment combines the advantage from the real prototype and virtual prototype which increases the efficiency and flexibility compared to the traditional method in the assembly work. In this way, it not only reduces the cost for the manufacturing industries, but also reduces the time and increase the accuracy during assembly process. The optimized solution also offers the opportunity for different manufacturing sectors to work in a same time. This innovative idea breaks down the sequential work process for collaboration in assembly and offers a platform which supports the features of overlaying and registering information on the workspace in a spatially meaningful way in AR. Moreover, it improves the human performance and relieves some of the workers' mental workload. The AR systems offer a real time collaboration and interaction between various enterprise units allowing a fast information exchange and letting the natural separation from designers and manufacturers to be overcome.

REFERENCES

Aliaga, D. G. (1994). Virtual and real object collisions in a merged environment. In *Virtual Reality Software Technology* [Singapore: World Scientific Publishing Co.]. *Proceedings of VRST*, *94*, 287–298.

Azuma, R. (1993). Tracking requirements for augmented reality. [Special issue on computer augmented environments]. *Communications of the ACM*, *36*(7), 50–52. doi:10.1145/159544.159581

Bolt, R. (1980). Put-That-There: Voice and gesture at the graphics interface. *Computer Graphics*, *14*(3), 262–270. doi:10.1145/965105.807503

Breen, D. E., Rose, E., & Whitaker, R. T. (1995). *Interactive Occlusion and Collision of Real and Virtual Objects in Augmented Reality*. (Technical report ECRC-95-02).

Chryssolouris, G., Pappas, M., Karabatsou, V., Mavrikios, D., & Alexopoulos, K. (2007, April). A Shared VE for Collaborative Product Development in Manufacturing Enterprises. In Li, W. D., Ong, S. K., Nee, A. Y. C., & McMahon, C. (Eds.), *Collaborative Product Design and Manufacturing Methodologies and Applications*. London: Springer-Verlag. doi:10.1007/978-1-84628-802-9_3

Constantinescu, C., Runde, C., Volkmann, J., Lalas, C., Sacco, M., Liu, D., Pavlopoulos, C., & Pappas, M. (2006). *DiFac D1 – Definition of a VR-based collaborative digital manufacturing environment*. DiFac Project (FP6-2005-IST-5-035079), v. 6.4.

Curtis, D., Mizell, D., Gruenbaum, P., & Janin, A. (1998). Several devils in the detail: Making an AR application work in the airplane factory. In *Proceedings of the International Workshop on Augmented Reality '98.*

Feiner, S., MacIntyre, B., & Seligmann, D. (1993). Knowledge-based augmented reality. [Special issue on computer augmented environments]. *Communications of the ACM*, *37*(6), 53–62. doi:10.1145/159544.159587

Liverani, A., Amati, G., & Caligiana, G. (2004). A CAD-Augmented Reality Integrated Environment for Assembly Sequence Check and Interactive Validation. *Concurrent Engineering*, *12*(1), 67–77. doi:10.1177/1063293X04042469

Mellor, J. P. (1995). *Enhanced reality visualization in a surgical environment. (Tech. Rep)*. MIT Artificial Intelligence Laboratory.

Metzger, P. J. (1993). *Adding reality to the virtual*. Paper presented at Virtual Reality Annual Symposium, WA, USA.

Molineros, J. M. (2002). *Computer Vision and Augmented Reality for Guiding Assembly*. PA: Department of Computer Science and Engineering in Pennsylvania State University, State College.

Neumann, U., & Cho, Y. (1996). A self-tracking augmented reality system. *In Proceedings of Virtual Reality Software and Technology (VRAIS96)* (pp. 109–115). Hong Kong.

Pang, Y., Nee, A. Y. C., Ong, M., Yuan, S. K., & Youcef-Toumi, K. (2006). *Assembly Automation*. Bradford, UK: Emerald Group Publishing Limited.

Pellom, B., & Kadri, H. (2003). Recent Improvements in the CU SONIC ASR System for Noisy Speech: The SPINE Task. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong.

Pope, A. R. (1994). *Model-based object recognition: a survey on recent research (Tech. Rep)*. California: University Berkeley.

Raghavan, V., Molineros, J., & Sharma, R. (1999). Interactive Evaluation of Assembly Sequences Using Augmented Reality. *IEEE Transactions on Robotics and Automation*, *15*, 435–449. doi:10.1109/70.768177

Sharma, R., & Molineros, J. (1997). *Computer vision-based augmented reality for guiding manual assembly*. Paper presented at the Virtual Reality Annual International Symposium, Atlanta, GA, USA.

Shin, D. H., & Dunston, P. S. (2008). *Evaluation of Augmented Reality in steel column inspection*. West Lafayette, IN: School of Civil Engineering, Purdue University.

Sims, D. (1994). New realities in aircraft design and manufacture. *IEEE Computer Graphics and Applications*, *14*(2), 91. doi:10.1109/38.267487

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proc. CHI* (pp. 73-80).

Tuceryan, M., Greer, D. S., Whitaker, R. T., Breen, D. E., Crampton, C., Rose, E., & Ahlers, K. H. (1995). *Calibration requirements and procedures for a monitor-based augmented reality system*. IEEE Trans. Visual. Computer. Graph.

Uenohara, M., & Kanade, T. (1995). *Vision-based object registration for real-time image overlay* (pp. 14–22). Computer Vision, Virtual Reality and Robotics in Medicine.

Urban, E. C. (1995). The information warrior. *IEEE Spectrum*, *32*(11), 66–70.

Vallino, J. R. (1998). *Interactive Augmented Reality*. University of Rochester.

Wiedenmaier, S., Oehme, O., Schmidt, L., & Luczak, H. (2001). Augmented reality for assembly process: An experimental evaluation. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality* (pp. 185–186). Los Alamitos, CA: IEEE Computer Society.

Wilson, R. H. (1995). Minimizing user queries in interactive assembly planning. *Robotics and Automation. IEEE Transactions on*, *11*, 308–312.

Yerkes, R. M., & Dodson, J. D. (1908). The Relationship of Strength of Stimulus to Rapidity of Habit Formation. *The Journal of Comparative Neurology and Psychology*, *18*, 459–482. doi:10.1002/cne.920180503

KEY TERMS AND DEFINITIONS

Augmented Reality: A field of computer research which deals with the combination of real-world and computer-generated data such as the computer generated scene overlaid on the real world.

Virtual Reality: A technology which allows the users to interact with a computer-simulated environment, whether that environment is a simulation of the real world or an imaginary world.

This work was previously published in Virtual Technologies for Business and Industrial Applications: Innovative and Synergistic Approaches, edited by N. Raghavendra Rao, pp. 105-117, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 98
E-Business/ICT and Carbon Emissions

Lan Yi
China University of Geosciences (Wuhan), China

ABSTRACT

The high-tech industry, its associated Information and Communication Technologies (ICT), and subsequent e-business and Internet applications, have had profound effects on our economic lives, social development and environment. This revolutionary change is not only appreciated by computer gurus, but also by private and public organisations. This chapter attempts to focus on the relationship between the e-business/ICT and carbon emission–one of the causes of global warming. Aspects to be considered include energy consumption and CO_2 emission associated with the Internet throughout its lifecycle-from manufacturing of Internet equipment/hardware to its usage and final disposal. On the other hand how Internet technology facilitates environment management and benefits industries via, for example Carbon Trading, will be discussed. The aim of this chapter is to provide a clarification and comprehensive picture of the carbon impact of ICT/e-business to private and public organisations as well as individuals, especially some "behind-the-scene" type of facts. Therefore environmental factors can be taken into account for more informed decision-making during business conduction.

INTRODUCTION

"In 1965, Gordon Moore, one of the visionary founders of Intel, predicted that the high-tech industry could double computing speed and power every two years. This vision, referred to as 'Moore's Law', has been realised" over and over again (SVTC, 2006a). This revolutionary change is not only appreciated by computer gurus, but also the whole society. Through one touch of a mobile phone, one can surf the Internet, send SMS and emails, or shoot 5-mega-pixels resolution photos and share them with the whole world, at a matter of seconds literally. What's more, these devices are improving constantly, becoming faster, smaller, cheaper, increasingly powerful and more

DOI: 10.4018/978-1-4666-1945-6.ch098

advanced. However, what shouldn't be ignored also is - they have shorter and shorter lifecycles. Taking pleasure in the revolution exultantly, few of us have noticed what sacrifices have been made to reach this achievement and what side effects are coming along.

Information and Communication Technologies, supported by the high-tech industry, and subsequent e-business and Internet applications, have made profound effects on our lives. The inevitable growing usage has marked it as digital economy era. On the other hand, with the recent Copenhagen Climate Change summit happening during a global recession, how to achieve a low carbon economy has also been on the top agenda of most governments. The further growth of these two trends seem unavoidable, however, are these two going to contribute or contradictory to each other? Study on the interaction between the environment and ICT is relatively new. And very few people really understand the full picture. The obvious links can be exemplified via, just to name a few, teleconferencing reducing carbon emission, online shopping replacing shopping malls, mp3 files substituting CDs, e-commerce deliveries causing additional transport pollution, extra packaging and increasing WEEE (Waste Electronic and Electrical Equipment) issues. Researchers have started to review these impacts from different angles using various methodologies. As can be expected from an emerging field of research, awareness and understanding of this topic is rather limited. Previous studies have presented mixed results.

This chapter attempts to focus on the interactive relationship between the Internet and the environment, energy consumption and carbon emission particularly. Further information regarding the broader linkages between the ICT/e-business and sustainable development (economic growth, environment protection and social development) can be found in the materials provided in the additional reading section at the end of this chapter.

The specific interaction between the e-business/ICT and carbon emission has been mostly looked at from the angle of "the Internet is associated with how much electricity". Rather than assessing the issue from this macroeconomic point of view, which is the most common approach, this chapter attempts to analyse it from a more detailed perspective, at the operational level. The aim is to provide a clearer and comprehensive picture of the carbon impact of e-business/ICT to private and public organisations as well as individuals, especially some "behind-the-scene" type of facts. Therefore they can take the environmental facts into account to make more informed decisions while they conduct business.

Aspects to be considered include (i) energy consumption and carbon emission associated with the Internet throughout its lifecycle, from manufacturing of Internet equipments/hardware to its usage and final disposal, and (ii) how Internet technology facilitates environmental management and benefits the energy industry via, for example Carbon Trading.

BACKGROUND

Mark P. Mills, one of the earliest to study the environmental impact of the Internet, claimed the energy consumed by Internet totalled about 8% of all US electricity use in 1998 and predicted it would grow to half of all electricity use in the next decade (Mills, 1999). These figures were claimed to be "significantly overestimated" by the Energy End-Use Forecasting and Market Assessment Group at Lawrence Berkeley National Laboratory (LBNL) (EUF, 2004). The debate reached an unprecedented level when the US House Subcommittee on National Economic Growth, Natural Resources, and Regulatory Affairs held hearings on "Kyoto and the Internet: the Energy Implications of the Digital Economy" on February 2nd 2000. At that hearing, Jay Hakes, Joseph Romm, and Mark Mills, the most authoritative and active figures in this research field testified. However, no unanimous conclusion was drawn as everybody looked at the problem from different angles and considered different system boundaries. Following this hearing, there were a few rounds of debates and rebuttals between Mills, LBNL etc, which received lots of attention in the media.

In 2005 LBNL published a final report for the PIER (Public Interest Energy Research) programme for the California Energy Commission (Masanet, Price, Can, Brown, & Worrell, 2005). Conclusion was made that despite the ubiquity and importance of PCs in industrialised economies, few data have been published - particularly data that characterise the life-cycle energy consumption and Greenhouse Gas (GHG) emissions of PC. This lack of data is due in large part to the extremely complex nature of the PC production system, which typically involves hundreds of different manufacturing processes, raw materials, and process chemicals, and is often dispersed among a vast global network of suppliers.

Though the results are rather inconsistent, these disputes aroused more interests than ever and triggered researchers and organisations to attempt to find the answer. Even though no conclusions have been made, the scientific value of this topic to researchers as well as the economic interest to industries and governments have been well exhibited. Yi and Thomas have provided a relatively comprehensive review of research on the environmental impact of e-business and ICT till 2007 (Yi & Thomas, 2007), including a review of related papers and thesis published, projects, comparisons of methodologies, conferences and symposiums. In this article, some of the most important development were mentioned and assessed, for example:

- Jokinen et al. examined the relationship between the information society and sustainable development on theoretical and conceptual levels (Jokinen, Malaska, & Kaivo-oja, 1998).
- Miller and Wilsdon published a paper entitled "Digital Futures — An Agenda for a Sustainable Digital economy" in 2001. This paper was one of the major products of the Digital Futures project, one of the most significant projects in this research field. In this paper, findings were summarised in the form of ten headline principles - what they call the ten 'dot-commandments' (Miller & Wilsdon, 2001).
- Matthews et al. compared the environmental and economic performances of traditional retailing and e-commerce logistic networks in the United States and Japan, using book retailing as a case study (Matthews, Williams, Tagami, & Hendrickson, 2002).
- In 2002, Romm published a journal paper entitled "The Internet and the new energy economy" (Romm, 2002), based on work first published in 1999 in a working paper (Romm, Rosenfeld, & Herrmann, 1999). These two papers have been the most frequently quoted and referred papers in this field.

With a glance of these titles, one can see previous efforts have been made from quite different perspectives. It is not feasible to conduct a direct comparison of the results as past work have chosen very different scopes in assessing the problem. Researches range from relatively small-scale case studies to very large pan-European projects. The range of the relationship also spans from strictly e-commerce to all encompassing digital technologies and electronic business platforms, and from a sub-sector of the environment, e.g. energy prospect, to general economic, social and environmental sustainable development. Nevertheless, past efforts have made significant contribution to the possible quantification of the issue in the future and set the background for this chapter.

The complexity of this topic originated from the complication of the Internet structure itself as well as the difficulty to define a clear-cut boundary for the input-output (economic as well as environmental) of the life cycle analysis. One simple network comprises large number of telecommunication equipments and terminals, from hub, router, gateway, server to PCs and peripherals, while every stage of the supply chain could involve energy consumption and CO_2 emission, from production to waste disposal. Furthermore, expertise needed to understand this issue incorporates a range of fields including computing, economics and environment science. Therefore the multi-disciplinary nature of the problem has also made the situation more complex.

Along the life cycle of e-business/ICT related physical existence, the next section of this chapter will provide (i) a review of how relevant e-business/ICT physical elements contribute to energy use and carbon emission in different stages and (ii) why e-business/ICT has contributed to data sharing and transactions on energy derivatives particularly the establishment of carbon trading platform.

ICT RESULTED CARBON EMISSION

This section attempts to break down the Internet associated energy consumption into the following levels along the life cycle of e-business/ICT related physical existence:

- The first section is related to the production which includes energy used in the production of Internet equipment/hardware, e.g. manufacturing of semi-conductors, PCs.
- The second section discusses energy related to and carbon emission resulted from shipment and distribution of components and equipment, e.g. overseas outsourcing.
- The third section presents energy consumption by the usage of hardware.
- The fourth section explains energy and carbon emission related to Internet and e-business applications, e.g. energy saved/used during usage of tele-conferencing and home-working.
- The last section reviews energy consumption and carbon emission related to the final disposal of the hardware.

Production

Along the life cycle of e-business/ICT related physical existence, this section discusses energy used in the production of Internet equipment/hardware, e.g. manufacturing of semi-conductors, PCs.

Researchers have considered different units in studies of computer related resource and energy consumption during the manufacturing stage, with reference to different mixes of processes in different regions. In general, it is agreed that computer related electronics manufacturing is a complicated process involving hundreds of natural resources and chemicals, requiring enormous amount of energy. Some compare the making of a single computer chip to boiling 10 gallons of water to make a cup of tea, as the Silicon Valley Toxics Coalition (SVTC) described. For example, the annual materials usage of a typical semiconductor facility (a semiconductor is an electronic component that is essential in many electrical devices from computers to mobile phones) includes: 832 million cubic feet of bulk gases, 5.72 million cubic feet of hazardous gases, 591 million gallons of deionised water, 5.2 million pounds of chemicals, and 8.8 million kilowatt hours of electrical power (SVTC, 2006b). This energy intensive process also results in large amount of carbon emission.

In 2002 Williams et al studied energy associated with manufacturing a 2-gram microchip in Japan (Williams, Ayres, & Heller, 2002). They stated the fossil fuel and chemical inputs to produce such a chip is estimated at 1600g and secondary materials used in production total 630 times the mass of the final product. This is an order of magnitude much larger than the factor of 1-2 for many other manufactured goods, such as automobiles and refrigerators.

The European Commission study (Atlantic-Consulting, 1998) broke down the assessment to components of a PC – control unit, keyboard, monitor etc and presented the results within the geographic boundary of the European Union as in Table 1 (unit: g CO_2-eqv). These results are however inconsistent. Furthermore data cannot easily be compared to others. Also, what shouldn't be ignored is that a PC or a PC component is just one part of an overall network. And the Internet is a vast collection of different networks that use certain common protocols and provide certain common services (Tanenbaum, 2003). So first of all, what physical elements (components/hardware/devices) are associated with the Internet that are contributing to carbon emission during their lifecycle?

The interconnected networks transmit data packets through guided transmission media (fibre-optic cables, twisted-pair copper wire or coaxial cable), or unguided media (e.g. wireless LAN or a digital satellite channel) (Kurose & Ross, 2003). The Internet's end systems normally include desktop computers (PCs, Macs, and UNIX-based workstations) and mobile computers (portable computers, PDAs and mobile phones with wireless Internet connections), as well as indirect servers (web servers, e-mail servers etc). An increasing number of alternative devices such as thin clients and household appliances, web TVs, set top boxes and digital cameras are also being attached to the Internet as end systems. Furthermore, there are devices which connect network segments between different layers, including repeaters, hubs, bridges, switches, routers and gateways. These switching devices are in common use but they operate in different layers in the OSI (ISO Open Systems Interconnection) reference model (Tanenbaum, 2003).

Also, some computer peripherals are likely to be attached for certain tasks. For example, in relation to e-business for the travelling sector, say a customer buying an e-ticket online, normally would be required to print the reference number/e-ticket or to save the webpage in a mobile. Therefore a printer or a mobile phone is considered to be part of the network.

Although different types of computer networks (e.g. LAN, WAN, MAN…) are available, the hardware and equipment involved are mostly a combination of elements mentioned above, which have carbon emission associated with each individual lifecycle. Nevertheless, the scope of elements involved far exceeds one type of equipment, e.g. PCs.

According to the US census, along with around 26.9 million electronic computers produced in 2006 in the US (US-Census-Bureau, 2007a), over $18,562 million (value of shipments) worth of wireless voice and data networking equipments were also produced, including for example, over $1,081 million worth of enterprise routers (US-Census-Bureau, 2007b). Thus if all Internet related hardware is considered, the contribution to global warming by manufacturing and using these devices will be significantly increased from earlier studies.

Calculation of these figures would be a confusing and daunting task. It is possible to collect data from a macroeconomics point of view, though it

Table 1. Emission of CO_2 equivalent from PC components

Components	Material production	Manufacture	Use	Distribution	Disposal	Total
Control unit	2.98E+04	5.32E+04	3.09E+02	1.68E+05	1.34E+03	2.49E+05
keyboard	3.21E+03	2.96E+03	4.35E+01	0.00E+00	6.07E+02	6.66E+03
monitor	1.64E+04	**7.69E+04**	**4.46E+02**	**2.80E+05**	**2.52E+03**	**3.75E+05**

will require data from different sectors including manufacturing of electronic computers, telecommunications equipment, semi-conductors, wiring devices and supplies etc. Nevertheless, when we sit in front of a PC either for personal or business purposes, we should bear in mind the long route taken by this unit and associated environmental impacts.

Shipment and Distribution of Components

Due to cost issues, technology availability and environmental regulations from different countries, e-business/ICT involved hardware are rarely locally manufactured. The material consumption and manufacturing process associated with the Internet hardware are rather energy intensive. On top of that, outsourcing and transportation of the hardware beyond geographical boundaries has made it more difficult to track the life cycle emissions. One PC could have components manufactured in hundreds of different locations and assembled elsewhere.

The reason for the migration and outsourcing of the hi-tech industry has gone beyond being a cost problem. After the discovery of massive groundwater contamination caused by toxic solvents from a computer chip factory in Silicon Valley of Santa Clara County, US in 1980s, hundreds of workers were found to develop cancer, reproductive problems, and other illnesses due to exposure to hazardous chemicals (SVTC, 2006c). The US Environmental Protection Agency had to deem 29 areas toxic Superfund sites, requiring priority environmental cleanup. As of 1996, 20 out of the 29 superfund sites were caused directly by the high-tech electronic industry and 5 were caused by related industries. Therefore, the electronic companies began outsourcing and moving their facilities overseas where labour is cheap and environmental protection laws are weak. "Today, behind any brand name electronic product a consumer buys off the shelf, there may have actually been 30 different companies involved with the making of that product" (SVTC, 2006c).

However, before considering the carbon emission involved in the transportation of computer related products, just how many computer related products are outsourced? Most well known electronic manufacturers such as Intel, AMD and Cisco are actually supported by many components makers worldwide. The quantities of outsourcing products can probably be reflected by the revenues of the overseas computer hardware markets. One of the most important outsourcing countries is undoubtedly China due to its cheap labour cost. China Development gateway, a Chinese NGO originally launched within the World Bank, stated that exports of computer products grew from $200 million in 1990 to $3.78 billion in 1995 (Chinagateway, 2004). In 1990 the Chinese computer industry had 191 hardware manufacturers and a total workforce of 100,000 people. By the end of 1995, a total of over 300,000 people worked for 1,000 hardware manufacturers. China's computer industry grew in output value from $921 million in 1990 to $6.37 billion in 1995. When moving such a large quantity of products across the world, no matter what transportation means is used, related carbon emissions are undoubtedly going to be significant. In addition, the market has grown considerably since 1995.

The calculation of freight related emission would require standardisation of indicators, measurements and units. However, just to give an idea of the scale of the problem: flying a fully loaded freight A727 (capacity 19 tonnes) just over 300 kilometres (less than the distance from the Midlands to Brussels) or running an articulated lorry over 9,000 kilometres could generate 10 tonnes of CO_2 equivalent emissions (DEFRA, 2000). So for export from China alone, how much carbon emissions would be involved in moving $3.78 billion worth of computers across the globe every year?

It's nearly impossible to calculate the amount of carbon emission coupled with shipment and outsourcing of Internet components, nonetheless, the above will give some idea of the magnitude of the problem.

Furthermore, Internet supported electronic commerce has facilitated online transactions around the world. The fact that a consumer can order almost anything from anywhere in the world at any time, which will most probably be packed and delivered individually the next day, changes the traditional shipment pattern and correspondingly deteriorates the environment and accelerates carbon emission. This will be discussed further in the Applications section.

Usage of Hardware

Once the hardware are shipped and distributed through warehouses and retailers, the transportation process temporarily stops. The hardware finally reach consumers and are ready to be used. This section provides a simple equation on how to calculate energy consumption during usage of the hardware.

According to the UK's Department for Environment Food and Rural Affairs (DEFRA), powering 86 PCs for a year uses 23,000 kWh of electricity and could generate 10 tonnes of CO_2 equivalent emissions (DEFRA, 2000). To calculate the electricity use of a particular computer or equipment is theoretically straight forward – just multiply its power with the usage time and sum the results from all the units involved. However, the process becomes more complicated as different makes, brands, and modes of operations (e.g. stand by, power save mode…) are involved. LBNL set an equation to calculate the average annual energy used by each piece of equipment as (Kawamoto, et al., 2001):

$$UEC_i = (SPM_i \times (PA_i \times HA_i + PL_i \times HL_i + PO_i \times HO_i) / 7) \times 365/10^3 + ((1\text{-}SPM_i) \times (PA_i \times (HA_i + HL_i) + PO_i \times HO_i) / 7) \times 365/10^3 + EPC_i$$

where

UECi = Unit Energy Consumption for equipment type i (kWh/year)

i = index for equipment type

PAi = Average active mod e power for equipment type i (W)

PLi = Average low- power mode power for equipment type i (W)

POi = Average off mode power for equipment type i (W)

HAi = Hours of operation in active m ode for equipment type i (hours/week)

HLi = Hours of operation in low-power mode for equipment type i (hours/week)

HOi = Hours of operation in off mode for equipment type i (hours/week)

SPMi = Power-management-enabled rate for equipment type i (%)

EPCi = Extra energy for printing or copying for equipment type i (kWh/year)6

365 = days per year

7 = days per week

The results from LBNL are based on a few assumptions including: the average lifetime of all types of network equipment is 4 years; Cisco products are selected as the representative model.

With regards to whether energy use and other environmental impacts in the manufacturing phase of PC equipment outweigh the usage phase, there have been controversial statements made in the past. Williams et al have claimed that for energy, the production phase becomes more relevant than the use phase, a reversal of the situation for products such as automobiles and many household appliances (Williams, Ayres, & Heller, 2002). Williams (2004) stated life cycle energy use of a computer is dominated by production (81%) as opposed to operation (19%). However the European Commission study claims the characterised data for global warming shows that the largest environmental contributions of PC come from the use stage where they are caused by the electricity consumption (4.47E+5, unit: gCO_2-eqv), which is over three times of the second largest contributor

- the manufacture stage (1.33E+5, unit: gCO_2-eqv) (Atlantic-Consulting, 1998).

Despite the disagreement among different studies, available research has mainly focused on the energy consumption of PC units, which are only one type of equipment running within the overall network. Other Internet equipment also consumes significant amounts of energy, e.g. the constant running of Internet routers, switches, servers etc. Unlike PCs, these equipments don't really have low-power, standby or power-saving mode. Due to their functions in the Internet structure, these equipments mostly have to be fully powered on 24/7, which means that the associated energy consumption is not negligible.

At the organisational level, it is relatively easy to calculate the electricity consumption and carbon emission associated with PC units. The figure could potential help an organisation's bottom line as well as its environmental performance.

Internet and E-Business Applications

The above section presented energy consumption related to the usage of e-business/ICT physical components. Calculation is rather straightforward. However unlike refrigerators or televisions, energy consumption of e-business/ICT equipments is not just self-resulted. Internet and e-business applications, for example, tele-conferencing, online shopping have significantly changed our life-style – in many ways have contributed to considerable energy saving and reduced carbon emissions compared with traditional life-style. Differentiated from previous section where the focus was on the primary impacts from e-business/ICT involved physical components, this section mainly focuses on the secondary effects created by the ongoing use and application of e-business/ICT.

Some claim that increasingly complex products, such as microchips, require additional secondary materials and energy to realise their lower entropy form, which offsets their dematerialisation contribution – the secondary materialisation proposition (Williams, Ayres, & Heller, 2002). However many still believe that Internet supported applications have been making a positive contribution to resources and energy saving in many areas and consequently to carbon reduction.

For example, the substitution of air travel with tele-conferencing has significantly reduced transport related carbon emissions. DEFRA stated that flying 14 international long haul trips or 110 short haul trips, or travelling 165,000 km on a train could generate 10 tonnes of CO_2 equivalent emissions (DEFRA, 2000). According to the Financial Times, the average transatlantic business flight uses 80,000 to 100,000 pounds of fossil fuel, which could easily be avoided by the use of teleconferencing (Cohen, 2000). Compared with these figures, the electricity consumption of tele-conferencing equipment and associated carbon emission are almost negligible.

Many companies have also started implementing Intranet based home-working schemes, which significantly reduce office space and related energy consumption. The construction, heating and lighting of buildings in the UK accounts for half of all energy consumed, half of carbon emissions and 40% of raw materials. But office space is only actually used for 22% of the average week (Goodman, Alakeson, & Jorgensen, 2004). Many organisations have seen huge savings through tele-working, e.g. British Telecom's flexi-working programme has enabled a 50% reduction in floor space since the mid 1990s; Sun Microsystems' iWork has reduced office space needs by 25% in four years (Goodman, Alakeson, & Jorgensen, 2004); At&T estimates a US $25 million per year saving in real estate through its virtual office programme (Allenby, 2001).

E-commerce related Internet shopping has also replaced physical shopping malls to a certain degree. In contrast to traditional business, some pure online businesses offer even more potential to eliminate retail space completely. C2C (consumer to consumer) e-commerce, for instance EBay, serving over 100 million users with more than 1.4 million items up for auction each day, is maintaining a virtual website rather than a huge retail store. Romm et al predicted the Internet economy could also render unnecessary as much as 3 billion square feet of buildings – some 5% of the US commercial floor space – which would save a considerable amount of construction-related energy (Romm, Rosenfeld, & Herrmann, 1999).

Other Internet facilitated applications, for instance, e-procurement has saved BT 5.8 tons of paper from 100+ page catalogues that were previously circulated and 1 ton of paper from 24,500 order pads that were previously printed (BT, 2003). Furthermore, the internet has made virtualisation of products and digitisation of information possible. Taking the music industry as an example, physical products such as CDs could easily be replaced with musical files downloadable from the Internet. Production as well as related warehouse, transport, etc. could be eliminated. The digitisation of standard business transactions is also changing office paper use (Yi & Thomas, 2006). Transformation of the earlier physical form of these products to virtual format has obviously made a saving in energy across their life-cycles.

Though on the negative side, the "Harry Potter" event – one of the milestone in the e-business history – was strongly criticised because of its environmental impact. On a single Saturday in July 2000, 100 airplanes and 9,000 trucks delivered more than 250,000 copies of 'Harry Potter and the Goblet of Fire' to Amazon.com customers all over the USA. This event had set a record for the online provision of goods in volume, but it probably also set a record for the quantities of empty shipping boxes and packaging that would wind up in landfills, let alone the energy used, transport pollution and carbon emission caused in the deal (Matthews, Hendrickson, & Lave, 2001).

The net impact on energy during the Internet application phase is yet to be further explored. However, in general the positive aspects are believed to outweigh the rebound effects. Realising and analysing each possible detailed aspect, where and how much energy is associated with, is a start.

Disposals of Hardware

During the life cycle of e-business/ICT hardware, the last stage would be the final disposal. Unlike other consumer electronic products such as refrigerators, consumers couldn't be busier throwing away outdated PCs and replacing them with newer versions. The increasingly shorter life cycle of these PCs and other related products (mobile phones, PDAs etc) has induced a unique market with unprecedented amount of demand, compared with products involved in previous major revolutions such as radios and TVs. The enormous appetite for PCs inevitably leads to a high PC disposal rates: for example, an estimated 10,000 PCs become obsolete in California alone each day (CAW, 2004; Masanet, Price, Can, Brown, & Worrell, 2005). On the other hand, despite the massive resources and energy used in manufacturing a relatively small-sized item, the hi-tech industry has designed the computers in such a way that it is more difficult for customers to upgrade than to replace the whole unit.

The EC Directive on Waste Electrical and Electronic Equipment (WEEE) (European-Commission, 2000), was agreed upon in January 2003 and entered into force in January 2007. The Directive aims to minimise the impact of electrical and electronic goods on the environment by increasing re-use and recycling, hence reducing the amount of WEEE going to landfill. IT and telecommunications equipment is among the 10 categories listed and accounts for a considerable portion of the total. WEEE now constitutes one of the fastest growing waste streams in the EU, with current estimates indicating a rate of growth three times that of average municipal waste.

In addition, the process of incineration or recycling of WEEE is not carbon emission free, especially taking into account this amount of WEEE produced. In a report produced by

Earthworks and Oxfam America, Farrell et al. stated that PCs alone contain a medley of metals, including gold, silver, aluminium, lead, copper, iron, zinc, and tin. Many of these materials could be salvaged at the end of the computer's life and recycled. But currently, most discarded computers are dumped in landfills or incinerated. Incineration of e-waste releases heavy metals and dioxin into the atmosphere. Some of the larger and older smelters have done extensive ecological damage. Worldwide, smelting adds about 142 million tons of sulphur dioxides to the atmosphere every year, which is 13 percent of total global emissions. And since smelters burn huge amounts of fuel, they also release substantial quantities of GHG, such as CO_2 and perfluorocarbons (Farrell, Sampat, Sarin, & Slack, 2004).

Furthermore, much of the e-waste produced in developed countries is actually shipped to developing countries to recycle. It is estimated that 50% to 80% of the "made-in-USA" e-waste is not recycled domestically but placed on container ships bound for destinations like China due to cheaper labour and lack of environmental standards in these countries (Puckett, et al., 2002). The process of transfer and transportation will cause another round of carbon emission.

CARBON TRADING PLATFORM

The previous section has presented the carbon emissions that Internet and related hardware manufacturing and usage have caused. On the other side, the Internet has enabled effective data sharing and ICT have been making positive contributions on energy issues in various industries. For example:

- The design of intelligent energy efficient products and systems has been facilitated, e.g. intelligent transport systems, environmental management systems, and energy management in housing.
- It can be conducted more effectively to monitor energy flow and other environmental indicators in the supply chain process: from production, distribution to disposal.
- Tools, software/application and platforms for energy trading, especially across countries, have been provided.

In particular, digital technologies supported carbon trading platforms have also created new markets online and is becoming an emerging and fast growing global phenomenon.

Background

Driven by the Kyoto Protocol signed in 1997, governments have committed to reduce carbon dioxide emissions by 20% by 2020. Carbon dioxide is one of the major GHG - the commonly known causes of global warming. Carbon emission has stretched its impact on our economies beyond being a simple environmental issue. On 30 October 2006, the UK government published a 579-page report on the economics of climate change by Sir Nicholas Stern. According to Stern (Stern, 2006), the overall costs and risks of climate change will be equivalent to losing 5% to 20% of GDP. In contrast, the costs of action – reducing GHG to avoid the worst impacts of climate change – can be limited to around 1% of global GDP each year. With no action, each tonne of carbon dioxide we emit will cause at least $85 (£45) of damage.

The British Government has introduced a tax on industrial energy as well as a cap and trade scheme. Companies are granted with an annual allowance of CO_2 emission (the cap). Whoever exceeds the cap will be fined; alternatively the company can buy allowances from other companies who have surplus. In short, industry societies now have to pay for their pollution, e.g. the price of 2005 CO_2 allowance rose to a record of €29.50 a ton on July 7th in the European CO_2 rights market (Levy, 2005). Due to immaturity of the system and policy, the price has fluctuated significantly during the first phase (2005-2007) as well as the recent recession.

The Market and Concept

Carbon is taken by the financial professionals as a form of energy derivatives, which is not a new concept. However, carbon does bear certain unique characters. Since European Commission and national governments still allocate free allowances to industries, only companies that exceed this free allowance needs to buy credits. The accuracy of this initial allocation has significant impact on the subsequent price. The price of carbon also fluctuates under influences of politics, country policies, other energy markets and so on. There are increasing number of regional, national and international systems for trading and transfer of GHG allowances and emission reduction credits. Hasselknippe reviewed a total of 47 systems for GHG trading and transfer included in the International Emissions Trading Association (IETA) Trading Schemes Database (Hasselknippe, 2003).

Various electronic platforms established domestically and internationally include the Climex Alliance (a pan-European market for trading Carbon Emissions Certificates within the EU's Emissions Trading Scheme), the Intercontinental Exchange (ICE), SENDECO2 (the Spanish CO_2 exchange expanding to Italy, Greece and Portugal) etc. The European Union Emission Trading Scheme (EU ETS), the largest multi-national GHG emissions trading scheme in the world and the main pillar of EU climate policy, commenced operation on 1st January 2005. All 27 member states of the European Union participated. In its first year, 362 million tonnes of CO_2 were traded on the market for a sum of €7.2 billion.

The international carbon market in 2005 transacted a total of 799 million tonnes of CO_2 equivalents worth approximately €9.4 billion (Point-Carbon, 2006). The carbon market grew in value to an estimated US$21.5 billion in the first three quarters of year 2006, more than doubling in value over the previous year (Capoor & Ambrosi, 2006). For such a large market to survive and the second phase (2008 ~ 2012) of the ETS scheme to proceed, technological support by ICT will play an important and growing role. Therefore how to integrate all these different systems presents a challenge. The communication and harmonisation of the systems as well as security control, can only be implemented probably with the support of the Internet and other Information and Communication Technologies.

Reduction of carbon emission has become an unavoidable issue for industries and governments. Carbon trading has been developed beyond the single purpose of protecting the environment with the involvement of international financial industry. With more participants and engagements, the mechanism and systems have become more and more complicated. Some even expect the system will further expand from energy industry to cover local government, retail sector, or even individual citizens. One day a carbon credit card for individuals to pay for electricity, gas and petrol might become reality (Sustainable-Development-Commision, 2006). No matter the future of the trading is a global central exchange or over-the-counter mechanism, it wouldn't succeed without the support of ICT – which helps to overcome the barrier of geographical boundaries and the challenge of large amount of data flowing around the globe.

The Case of EU

PriceWaterHouseCooper has reported the state of art of potential expanded use of IT in EU-ETS (PriceWaterHouseCoopers, 2006). It summarised that in the current EU-ETS, IT is restricted to support monitoring, reporting and registry activities. Verifiers and competent authorities use non-standardised software for internal purposes only. For example in the case of the UK, in the Verification Reference Model of EU ETS, IT is used as (i) a template for reporting, (ii) a template for verifiers and (iii) the home developed registry. No software application is currently used in the monitoring stage and the electronic media for reporting is not currently web enabled. The vision of British regulators is to have a web enabled reporting system with different levels of access for regulators, operators, verifiers and the public and a standardised system to be shared across the EU.

The national registry in the UK is accessed via the Internet and has been developed using DEFRA's technical design standards in Microsoft.NET 1.1 and SQL Server. By using web services the registry is able to communicate with the EU Commission's Community Independent Transaction Log (CITL). Currently the communication is secure and processed as real time transactions. The UN Data Exchange Standards (DES) specifies the use of TCP/IP connections using encrypted messages over the Internet in order to protect communications from modification or interception in transit (PriceWaterHouseCoopers, 2006).

During the trading phase, DEFRA (UK) and the French Registry developers, Caisse des ustus et consignations (CDC), are working together to enable trading platforms to interface with the registry systems. DEFRA hopes that the creation of international data exchange standards will facilitate the development of electronic interfaces by end users allowing trading platforms and electronic accounting systems to communicate automatically with Registries (PriceWaterHouseCoopers, 2006).

In order to accommodate growing usage and future demands, transparency, reliability, speed, capacity, adaptabilities, upgradeability and security of these online platforms are critical for the trading to operate effectively and efficiently. Development of a sound IT infrastructure and harmonised interfacing between different systems will be fundamental to the overall success of the programme.

CONCLUSION

ICT and related Internet usages have made profound effects on our economic lives, social development and environment. Research on the environmental impacts of the ICT is relatively new. This chapter attempted to explore the interaction between Internet technologies, energy consumption and carbon emission. The objective is to give a detailed, however not exhaustive, summary of various stages of how the Internet and e-business contribute to the energy issues. Private and public organisations can therefore more clearly understand the environmental impact of their usage of related technologies in their business.

As expected, research that has made a significant difference is limited and is far from finding any definitive conclusions or solutions. Also the depth of research is not enough. However on the positive side, significant interest has been provoked. Awareness of issues involved has also been raised not only with academics but also with policy makers and industries. Some results have paved the way for future research in this field.

With reference to the manufacturing of Internet hardware and equipment, past research has considered different system boundaries with different mixes of processes in different regions. Nevertheless it is agreed that computer related electronics manufacturing is a complicated but clearly energy intensive process. Outsourcing and transportation beyond geographical boundary has made it difficult to track the life cycle of Internet hardware. The usage stage is claimed to involve large amounts of energy due to the constant running of networking hardware but relatively easy for individuals to calculate. Fortunately subsequent Internet supported applications have been making a positive contribution to resources and energy saving in many areas and consequently to carbon reduction. These include home-working, tele-conferencing and dematerialisation of physical products. Finally, the creation of e-waste and its transfer to developing countries has however caused another round of negative ecological impacts.

Additionally, the chapter has also looked at the importance of ICT in facilitating the emerging carbon trading systems. In order to accommodate growing usage and future demand and for the trading to operate effectively and efficiently, the systems have to be transparent, reliable, adaptable, upgradeable, and secure. These critical requirements will all largely depend on advanced ICT and web-enabled technologies.

Understanding of the relationship between the Internet and carbon emission has demonstrated substantial impact on environmental monitoring and protection. Quantitative standardised methodologies are currently under development and will be beneficial in realising the full picture. Nonetheless, presenting a full picture of the interaction hopefully can help both private and public organisations see the facts behind the scene, and therefore manage their purchase of IT equipments and integrate IT into business more sensibly.

SOLUTIONS AND RECOMMENDATIONS

This chapter essentially tries to provide a picture of and reference on how ICT and e-business interacts with the environment, particularly the carbon and energy side along the life cycle of the ICT physical elements. A literature review and survey of the current status is expected to facilitate future research and further understanding.

It is recommended that private and public organisations devote appropriate amount of time and resources to measure the environmental impact such as GHG emissions associated with their general business operations – including ICT and e-business usage. Many industry sectors including mining, utilities, and infrastructure have seen the business value and opportunities of conducting business in an environmentally sound and socially responsible manner.

Organisations are suggested to take on the positive contribution from ICT sector and e-business platforms and to carry out business more sustainably. This includes not only optimising internal operations (such as management of electricity, water, gas use, transportation and waste), but also making the choice of having more sustainable partners. Examples of achieving a greener business might start with a business manager taking a video-conferencing instead of flying across Atlantic, or simply mean reusing or recycling an old PC. Different organisation/company may have different operational strategies, but raised awareness and taking account the environmental factor into business decisions, will only help the planet as well as individual's bottom line.

FUTURE RESEARCH DIRECTIONS

Current dominant approach of analysing the problem is either a micro-level case study approach or a macro-level statistical approach. Traditional assessment approaches are found to be insufficient to accommodate the digital technology revolution and cannot accommodate the challenge of measuring the impacts of ICT on environmental sustainability. New innovative methods need to be created to fill this gap. It is suggested that a more predictive and empirical model, which can be applied within a sector of society, should be more beneficial in the long term. This approach should help simulate potential impacts resulting from changes of indicators, so that positive effects can be promoted and negative ones alleviated proactively, rather than knowing and accepting outcomes passively (Yi & Thomas, 2007).

The challenge of any research does not stop at recognising the problem, but to know: how to solve the problem, alternative solutions with corresponding risks and benefits, and how to choose the best solution. It is not enough to know ICT/e-business has been changing our daily life, economy, transport, air, water, forests, etc., it is not even enough to understand how it is changing everything, ultimately an approach which can influence the behaviour, say of a private organisation for example, is needed.

REFERENCES

Atlantic-Consulting. (1998). *LCA study of the product group personal computers in the EU Eco-label scheme.* Retrieved from http://ec.europa.eu/environment/ecolabel/pdf/personal_computers/lcastudy_pc_1998.pdf

BT. (2003). *E-business and the environment. Better world–our commitment to society*. Report retrieved from www.groupbt.com/betterworld

Capoor, K., & Ambrosi, P. (2006). *State And trends of the carbon market, 2006*. Washington, DC: IETA and the World Bank.

CAW. (2004). *E-waste background. Californians Against Waste.* Retrieved from www.cawrecycles.org/Ewaste/background%20e-waste.html

Chinagateway. (2004). *China's electronics industry.* PRC Ministry of Posts and Telecommunications. Retrieved from http://en.chinagate.com.cn/english/1474.htm

Cohen, A. (2000). Catching the green express from euston. *Financial Times*.

DEFRA. (2000). *Environmental reporting–guidelines for company reporting on greenhouse gas emissions*. Retrieved from http://www.defra.gov.uk/environment/business/reporting/pdf/envrpgas-annexes.pdf

EUF. (2004). Information *Technology and resource use, energy end-use forecasting at Lawrence Berkeley National Laboratory*. Retrieved from http://enduse.lbl.gov/Projects/InfoTech.html

European-Commission. (2000). *Directive on waste electrical and electronic equipment and directive on the restriction of the use of certain hazardous substances in electrical and electronic equipment.* Explanatory memorandum. Brussels.

Farrell, L., Sampat, P., Sarin, R., & Slack, K. (2004). *Dirty metals-mining, communities and the environment*. Earthworks and Oxfam America.

Goodman, J., Alakeson, V. & Jorgensen, B. (2004). *Encouraging green telework*: *Forum for the future*.

Hasselknippe, H. (2003). Systems for carbon trading: An overview. *Climate Policy*, *3*(2), S43–S57. doi:10.1016/j.clipol.2003.09.014

Jokinen, P., Malaska, P., & Kaivo-oja, J. (1998). The environment in an information society: A transition stage towards more sustainable development? *Futures*, *30*(6), 485–498. doi:10.1016/S0016-3287(98)00054-8

Kawamoto, K., Koomey, J. G., Nordman, B., Brown, R. E., Piette, M. A., & Ting, M. (2001). *Electricity used by office equipment and network equipment in the U.S.: Detailed report and appendices*. Orlando: Lawrence Berkeley National Laboratory.

Kurose, J. F., & Ross, K. W. (2003). *Computer networking: A top-down approach featuring the Internet* (2nd ed.). Pearson Education, Inc.

Levy, A. (2005). *Making pollution pay*. Bloomberg Markets.

Masanet, E., Price, L., Can, S. R., Brown, R., & Worrell, E. (2005). *Optimization of product life cycles to reduce greenhouse gas emissions in California*. California Energy Commission, Public Interest Energy Research Program.

Matthews, H. S., Hendrickson, C. T., & Lave, L. (2001). How much did Harry Potter cost? E-commerce is growing and environmentalists should be happy. Or should they? Retrieved from http://www.oecdobserver.org/news/fullstory.php/aid/415/How_much_did_Harry_Potter_cost_.html

Matthews, H. S., Williams, E., Tagami, T., & Hendrickson, C. T. (2002). Energy implications of online book retailing in the United States and Japan. *Environmental Impact Assessment Review*, *22*(5), 493–507. doi:10.1016/S0195-9255(02)00024-0

Miller, P., & Wilsdon, J. (2001). Digital futures-an agenda for a sustainable digital economy. *Corporate Environmental Strategy*, *8*(3), 275–280. doi:10.1016/S1066-7938(01)00116-6

Mills, M. P. (1999). *The Internet begins with coal*. Greening Earth Society.

Point-Carbon. (2006). [*-towards a truly global market.*]. *Carbon*, 2006.

PriceWaterHouseCooper. (2006). *Information Technology enhances consistency, transparancy efficiency in EU-ETS.*

Puckett, J., Byster, L., Westervelt, S., Gutierrez, R., Davis, S., & Hussain, A. (2002). *Exporting harm-the high-tech trashing of Asia. The Basel Action Network (BAN), Silicon Valley Toxics Coalition*. SVTC.

Romm, J. (2002). The Internet and the new energy economy. *Resources, Conservation and Recycling*, *36*(3), 197–210. doi:10.1016/S0921-3449(02)00084-8

Romm, J., Rosenfeld, A., & Herrmann, S. (1999). *The Internet economy and global warming: A scenario of the impact of e-commerce on energy and the environment. Center for Energy and Climate Solutions (CECS), Global Environment & Technology Foundation*. GETF.

Stern, N. (2006). *The economics of climate change: The Stern Review*. Retrieved from http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm

Sustainable-Development-Commision. (2006). Carbon trading for individuals? Retrieved from http://www.sd-commission.org.uk/pages/carbon-trading.html

SVTC. (2006a). Electronics industry-a dazzling industry with a dark side. Retrieved February 15, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_electronic_industry_overview

SVTC. (2006b). Electronics manufacturing-a deadly process requiring thousands of toxic chemicals Retrieved February 15, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_electronics_production

SVTC. (2006c). High-tech history: Chronicling the evolution of the high-tech industry. Retrieved February 15, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_hitechhistory

Tanenbaum, A. S. (2003). *Computer networks* (4th ed.). New Jersey: Pearson Education, Inc.

US-Census-Bureau. (2007a). *Computers and peripheral equipment: 2006 summary (Current industrial reports)*. Retrieved from http://www.census.gov/

US-Census-Bureau. (2007b). *Telecommunications: 2006 (Current industrial reports)*. Retrieved from http://www.census.gov/

Williams, E. (2004). Energy intensity of computer manufacturing: Hybrid analysis combining process and economic input-output methods. *Environmental Science & Technology*, *28*(22), 6166–6174. doi:10.1021/es035152j

Williams, E. D., Ayres, R. U., & Heller, M. (2002). The 1.7 kilogram microchip and material use in the production of semiconductor devices. *Environmental Science & Technology*, *36*(24), 5504–5510. doi:10.1021/es025643o

Yi, L., & Thomas, H. R. (2006). E-business and sustainable development. *International Journal of Environment and Sustainable Development*, *5*(3), 262–274. doi:10.1504/IJESD.2006.010896

Yi, L., & Thomas, H. R. (2007). A review of research on the environmental impact of e-business and ICT. *Journal of Environment International*, *33*(6), 841–849. doi:10.1016/j.envint.2007.03.015

ADDITIONAL READING

Abukhader, S. M., & JoÈnson, G. (2003). The environmental implications of electronic commerce: a critical review and framework for future investigation. *Management of Environmental Quality: An International Journal*, *14*(4), 460–476. doi:10.1108/14777830310488685

Arnfalk, P. (2004). The future impact of ICT on environmental sustainability: Evaluation and Recommendations (Fifth Interim Report) Lund: Institute for Prospective Technological Studies (IPTS).

Atkyns, R., Blazek, M., & Roitz, J., & AT&T. (2002). Measurement of environmental impacts of telework adoption amidst change in complex organizations: AT&T survey methodology and results. *Resources, Conservation and Recycling*, *36*, 267–285. doi:10.1016/S0921-3449(02)00082-4

Berkhout, F., & Hertin, J. (2004). De-materialising and re-materialising: digital technologies and the environment. *Futures*, *36*(8), 903–920. doi:10.1016/j.futures.2004.01.003

Caudill, R., Luo, Y., Wirojanagud, P., & Zhou, M. (2001). A life cycle environmental study of the impact of e-commerce on electronic products. Paper presented at the Proceedings of the Institute of Electrical and Electronics Engineering (IIIEE) International Symposium on Electronics and the Environment, USA.

Durieu, X. (2003). How Europe's retail sector helps promote sustainable production and consumption. *UNEP Industry and Environment*, *26*(1), 7–9.

Editorial. (2003). Shopping for a better world: sustainability and retailing. UNEP Industry and Environment, 26(1), 3-6.

Erdmann, L., & Behrendt, S. (2003). The future impact of ICT on environmental sustainability: Script (Second Interim Report). Berlin: Institute for Prospective Technological Studies (IPTS).

Erdmann, L., & Würtenberger, F. (2003). The future impact of ICT on environmental sustainability: Identification and global description of economic sectors (First Interim Report). Berlin: Institute for Prospective Technological Studies (IPTS).

Fichter, K. (2003). E-commerce - Sorting Out the Enviornmental Consequences. *Journal of Industrial Ecology*, *6*(2), 25–41. doi:10.1162/108819802763471762

Goodman, J., & Alakeson, V. (2003). The future impact of ICT on environmental sustainability: Scenarios (Third Interim Report). London: Institute for Prospective Technological Studies (IPTS).

Hesse, M. (2002). Shipping news: the implications of electronic commerce for logistics and freight transport. *Resources, Conservation and Recycling*, *36*(3), 211–240. doi:10.1016/S0921-3449(02)00083-6

Hilty, L. M., Arnfalk, P., Erdmann, L., Goodman, J., Lehmann, M., & Wager, P. A. (2006). The relevance of information and communication technologies for environmental sustainability - A prospective simulation study. *Environmental Modelling & Software*, *21*, 1618–1629. doi:10.1016/j.envsoft.2006.05.007

Hurst, E. (2001). E-commerce and the Environment. Unpublished MSc Thesis, University of London, London.

John, A. (2003). Information Technology and U.S. Energy Consumption: Energy Hog, Productivity Tool, or Both? *Journal of Industrial Ecology*, *6*(2), 13–24.

Macauley, M., Palmer, K., & Shih, J.-S. (2003). Dealing with electronic waste: modeling the costs and environmental benefits of computer monitor disposal. *Journal of Environmental Management*, *68*(1), 13–22.

Matthews, H. S., Hendrickson, C. T., & Lave, L. (2001). How much did Harry Potter cost? - E-commerce is growing and environmentalists should be happy. Or should they? [Electronic Version]. Retrieved 15/02/2010 from http://www.oecdobserver.org/news/fullstory.php/aid/415/How_much_did_Harry_Potter_cost_.html.

Matthews, H. S., Williams, E., Tagami, T., & Hendrickson, C. T. (2002). Energy implications of online book retailing in the United States and Japan. *Environmental Impact Assessment Review*, *22*(5), 493–507. doi:10.1016/S0195-9255(02)00024-0

Miller, P., & Wilsdon, J. (2001). Digital Futures - An Agenda for a Sustainable Digital Economy. *Corporate Environmental Strategy*, *8*(3), 275–280. doi:10.1016/S1066-7938(01)00116-6

Mills, M. P. (1999). The Internet Begins with Coal: Greening Earth Society.

Park, J., & Roome, N. (Eds.). (2002). The Ecology of the New Economy: Sustainable Transformation of Global Information, Communications and Electronics Industries: Greenleaf Publishing.

Plepys, A. (2002). The grey side of ICT. *Environmental Impact Assessment Review*, *22*, 509–523. doi:10.1016/S0195-9255(02)00025-2

Point-Carbon. (2006). [- towards a truly global market.]. *Carbon*, 2006.

PriceWaterHouseCoopers. (2006). Information Technology enhances consistency, transparancy efficiency in EU-ETS.

Reijnders, L., & Hoogeveen, M. J. (2001). Energy effects associated with e-commerce: A case-study concerning online sales of personal computers in The Netherlands. *Journal of Environmental Management*, *62*, 317–321. doi:10.1006/jema.2001.0440

Rejeski, D. (2003). E-Commerce, the Internet, and the Environment. *Journal of Industrial Ecology*, *6*(2), 1–3. doi:10.1162/108819802763471717

Romm, J. (2002). The Internet and the new energy economy. *Resources, Conservation and Recycling*, *36*(3), 197–210. doi:10.1016/S0921-3449(02)00084-8

Romm, J., Rosenfeld, A., & Herrmann, S. (1999). *The Internet Economy and Global Warming: A Scenario of the Impact of E-commerce on Energy and the Environment Center for Energy and Climate Solutions (CECS), Global Environment & Technology Foundation*. GETF.

Roome, N., & Park, J. (2000). Global Sustainability and Information Economy: Old Challenges, New Perspectives. *Greener Management International*, *32*, 24–32.

Siikavirta, H., Punakivi, M., Kärkkäinen, M., & Linnanen, L. (2003). Effects of E-Commerce on Greenhouse Gas Emissions: A Case Study of Grocery Home Delivery in Finland. *Journal of Industrial Ecology*, *6*(2), 83–97. doi:10.1162/108819802763471807

Sioshansi, F. (2000). E-commerce and the Energy Sector: The Pioneers May Not Get It Right; The Procrastinators Are Likely to Become History. *The Electricity Journal*, *13*(5), 42–49. doi:10.1016/S1040-6190(00)00117-2

Stern, N. (2006). The Economics of Climate Change: The Stern Review. Retrieved. from http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm.

Sui, D. Z., & Rejeski, D. W. (2002). Environmental Impacts of the Emerging Digital Economy: The E-for-Environment E-Commerce? *Environmental Management*, *29*(2), 155–163. doi:10.1007/s00267-001-0027-X

SVTC. (2003). Just say no to e-waste: background document on hazards and waste from omputers [Electronic Version]. Retrieved 15/02/2010 from http://www.svtc.org/cleancc/pubs/sayno.htm.

SVTC. (2006a). Electronics Industry - A Dazzling Industry With a Dark Side. Retrieved 15/02/2010, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_electronic_industry_overview

SVTC. (2006b). Electronics Lifecycle - A Wake of Unintended Collateral Damage from Cradle to Coffin Retrieved 15/02/2010, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_lifecycle_analysis

SVTC. (2006c). Electronics Manufacturing - A Deadly Process Requiring Thousands of Toxic Chemicals Retrieved 15/02/2010, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_electronics_production

SVTC. (2006d). High-tech History: Chronicling the Evolution of the High-Tech Industry Retrieved 15/02/2010, 2010, from http://svtc.etoxics.org/site/PageServer?pagename=svtc_hitechhistory

Tongia, R., Subrahmanian, E., & Arunachalam, V. S. (2004). *Information and Communications Technology for Sustainable Development: Defining a Global Research Agenda (No. 81-7764-839-X)*. Bangalore, India: Carnegie Mellon University, Pittersburgh & Indian Institute of Science, Banglore.

Tongia, R., Subrahmanian, E., & Arunachalam, V. S. (2005). *Information and Communications Technology for Sustainable Development: Defining a Global Research Agenda*. Bangalore: Allied Publishers.

Velásquez, M., Ahmad, A.-R., & Michael, B. (2009). State-of-the-Art in E-Commerce Carbon Footprinting. *Journal of Internet Banking and Commerce*, *14*(3), 1–21.

Williams, E. (2004). Energy intensity of computer manufacturing: hybrid analysis combining process and economic input-output methods. *Environmental Science & Technology*, *28*(22), 6166–6174. doi:10.1021/es035152j

Wilsdon, J. (Ed.). (2001). *Digital Futures: Living in a Dot-Com World*. London: Earthscan.

Wilsdon, J., & Miller, P. (2001). Digital Futures: an agenda for a sustainable digital economy: Forum for the Future.

Yi, L., & Thomas, H. R. (2006). E-business and Sustainable Development. *International Journal of Environment and Sustainable Development*, *5*(3), 262–274. doi:10.1504/IJESD.2006.010896

Yi, L., & Thomas, H. R. (2007). A review of research on the environmental impact of e-business and ICT. *Journal of Environment International*, *33*(6), 841–849. doi:10.1016/j.envint.2007.03.015

Yi, L., & Thomas, H. R. (2009). A Decision Support System for the Environmental Impact of ICT and E-business. [IJITDM]. *International Journal of Information Technology & Decision Making*, *8*(2). doi:10.1142/S0219622009003387

KEY TERMS AND DEFINITIONS

Carbon Market: Refers to the Emission Trading market, is one of the three mechanisms designed in the Kyoto Protocol. The other two are namely Clean Development Mechanism and Joint Implementation. The term is also called the cap and trade market, since normally a government sets the limit / cap on the amount of Greenhouse Gas (GHG) that can be emitted. Essentially it is an approach designed to control GHG emission through economic mechanism. Companies are issued with specific permit to emit certain amount of GHG. For those who exceeds the allowance, they either have to pay a fine, or they can buy the allowances from another company who emits less than its assigned allowances. The transfer of the allowances therefore becomes a trade. Transfers and acquisitions of these allowances are tracked and recorded through the registry systems under the Kyoto Protocol. An international transaction log ensures secure transfer of emission reduction units between countries.

DEFRA: The UK government department responsible for policy and regulations on the environment, food and rural affairs.

Electronic Business (E-Business): Is the application of Information and Communication Technologies in any essential activities of a business throughout its (external and internal) supply chain. The term can be confused with the other concept "e-commerce", which is a subset of e-business and normally only includes conducting business via online platforms.

Greenhouse Gases: The atmospheric gases that contribute to the greenhouse effect, including water vapor, carbon dioxide, nitrous oxide, methane etc. These gases are believed to be able to trap heat therefore affect the temperature of the earth.

Information and Communication Technologies (ICT): Is a term used to encompass computing, telecommunication, digital technologies that intend to facilitate the processing, analysing, transfering, exchanging, distribution of information.

Kyoto Protocol: Is an international agreement linked to the United Nations Framework Convention on Climate Change. The Protocol was initially adopted on 11 December 1997 in Kyoto, Japan and entered into force on 16 February 2005. The major feature of the Kyoto Protocol is that it sets binding targets for 37 industrialised countries and the European community for reducing Greenhouse gas (GHG) emissions.

Open System Interconnection (OSI): Reference Model: Is an industry wide protocol and international agreement for layered communications and computer network. It divides network architecture into seven layers as: the Application, Presentation, Session, Transport, Network, Data Link, and Physical Layers. It is therefore often referred to as the OSI Seven Layer Model.

Peripheral: Is a device that attached to a host computer. It normally does not constitute as the core component but it extends the functions of the main computer. Examples include printer, scanner, speaker, webcam.

Silicon Valley Toxics Coalition (SVTC): Created in 1982, Silicon Valley Toxics Coalition is a non-profit organisation engaged in research, advocacy and grassroots organising to promote human health and environmental justice in response to the rapid growth of the high-tech industry. It is based in San Jose, California, the United States.

Superfund Sites: According to United States Environment Protection Agency (EPA), a Superfund site is an uncontrolled or abandoned place where hazardous waste is located, possibly affecting local ecosystems or people. It is also the name of the fund established by the Comprehensive Environmental Response, Compensation and Liability Act of 1980, as amended (CERCLA statute, CERCLA overview). This law was enacted in the wake of the discovery of toxic waste dumps such as Love Canal and Times Beach in the 1970s. It allows the EPA to clean up such sites and to compel responsible parties to perform cleanups or reimburse the government for EPA-lead cleanups.

This work was previously published in Impact of E-Business Technologies on Public and Private Organizations: Industry Comparisons and Perspectives, edited by Ozlem Bak and Nola Stair, pp. 214-232, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 99
Building for the Future:
Systems Implementation in a Construction Organization

Hafez Salleh
University of Malaya, Malaysia

Eric Lou
University of Salford, UK

EXECUTIVE SUMMARY

This chapter provides the IT readiness assessment for before and after scenarios of IT systems implementation in a construction consultancy company providing multi-disciplinary services for the construction industry throughout the United Kingdom. The services offered include building surveying, quantity surveying, project management, civil and structural engineering design, and mechanical and electrical engineering design, among others. On application of the maturity model it was found that the overall processes for managing information are improving since the introduction of the new IT system. Prior to the project, the development of IT/IS was driven to perform daily work tasks that required the company to run a business. The new systems has streamlined the organization-wide communication, which the previous system did not have the capability to do, and to reduce cost for document reproduction. The level of IT skills prior to the project was relatively low; the introduction of the new system has helped the company to increase their staff's IT skills.

BACKGROUND AND HISTORY

Organization B is a construction consultancy company providing multi-disciplinary services for the construction industry throughout the United Kingdom. The services offered include building surveying, quantity surveying, project management, civil and structural engineering design and mechanical and electrical engineering design, among others. Organization BBBBB was established in 1941 and operates from their offices in four different cities. The organization turned into a Limited Liability Partnership (LLP) in April 2006, with an annual estimated turnover of £12 million. Organization BBBBB employs 220 staff across 4 offices, of which 120 staff are

DOI: 10.4018/978-1-4666-1945-6.ch099

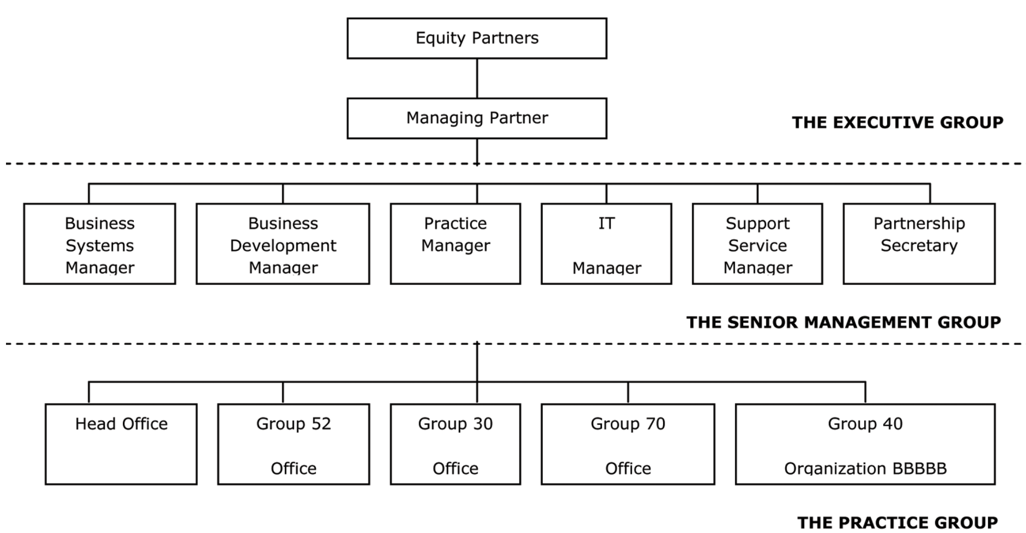
located at the Head Quarters of the company. No specific department exists in their organizational structure. Instead, the organization operates in groups, but not strictly by discipline. For example, one group consists of multiple disciplines, and anyone can be a group leader. The disciplines are as follows: Building Surveyor, Quantity Surveyor, Project Manager, Employers Agent, Architect, CAD, Mechanical and Electrical Engineers, Civil and Structural Engineers, Planning Supervisors. There are three layers of management within Organization BBBBB's organization structure; The Executive Group, The Senior Management Group and The Practice Group. Organization BBBBB's organization structure is shown in Figure 1.

Sequence of Events

In November 2004, the management of the company discussed the need to replace their existing Database Management System (DMS), which had reached the limit of its development potential. In doing this, they had to recognize the need for a coherent approach to their IT, and the need to bring all systems under one control.

In 2004, the management wanted to have access to all parts of the knowledge system through one front end. Further discussions established that they wanted to include a Knowledge Management System (KMS), and added Human Resource (HR) functionality to the proposed system. The organizational business system managers suggested what the practice really needed was a single piece of software that had the capability to manage all the functions within the organization. The management agreed, and asked for a time frame. The project team was then established and comprised of a business system manager as head, and IT manager, a database programmer and a business development manager; and started to evaluate the products available in the market. During the product evaluation, the project team identified products which could do more than just DMS functions, and in fact provide them with a Client Relation Manager (CRM) which is essentially an advanced address database with functions that make the information both accurate and usable, and also an extranet function which makes all documents retrievable from remote locations and an intranet for internal information distribution. The whole function combined to provide a sophisticated Knowledge Management System (KMS). This is what they thought many people had been requesting for some time. The project

Figure 1. Organization BBBBB's organizational structure



team identified the preferred vendors, and five vendors were invited to present their products. One of the products that impressed the project team was the product from Vendor B, and this vendor was invited to present one of their products called Workspace ™. The presentation was attended by the managing partner, business system manager, IT manager and the database programmer. In addition to Workspace ™, the project team also reviewed other products by attending seminars, conferences, and seeking views from the companies who had implemented similar systems.

The product evaluation process took up to a year (then 2005). Finally, the project team decided to select the product from Vendor B as it met the specification of the required system for the company to bring all systems under one umbrella. A payment was made to Vendor B of about £200,000 and internal development costs were going to be added to this later. After signing the contract with the vendor, the business systems manager came out with a methodology (programme) of software development, including the necessary adjustments to the software and the time frame. Both these were represented in terms of business processes.

- **Initial phase**
    - looking at the corporate appearance of software
    - In house staff designed this corporate appearance
    - Looking at the extra functions to be added to the software
    - Within three months after that the management decided there was an urgent need to be looking at a human resources (HR) package
- **Second phase**
    - Started to evaluate a few human resource (HR) products to include human HR functionality to the workspace.
    - Started to load data into the HR product
    - Human resource managers complained the software was not functioning as expected.
    - The vendor who was supplying the HR product was very unhelpful, and imposed extra charges on every extra hour they spent on it which totally contrasted with Vendor B.
    - Vendor B put in a lot of time to put the system in order without extra charges
- **Third phase**
    - Terminated the contract with the vendor that supplied the HR product
    - Asked Vendor B to incorporate the HR function to their product
    - Started the development process
    - Steering committee (SC) formed to represent all functions and levels in the practice
    - Gave a presentation to SC about what the team had been doing with Vendor B so far, and sought their comments and agreement.
    - The SC were asked to look at and test the system, but it did not go well
    - There are several stages of testing
        - The in-house staff renamed the product as 'System Y' (for corporate appearance)
        - Selected staff tested the system – open, enter and save documents
        - It started in the Winchester office
        - Management expressed concern over time delays and fears of the system crashing if large numbers of users used it simultaneously
        - The business development group also participated in testing the system and commented on it.
        - Training managers went to each office branch and provided one

to one training to each member of staff

- All the members were trained to create and save documents in System Y. This is a basic level of training, and staff were certified to use the basic operations of System Y.
- Some of the partners said they wanted to work remotely, for example at home. The business system manager suggested it would be possible if a scan function could be added to the system in order to review documents off site.

- **Fourth phase**
  - This phase focused on the issue of how to maximize the capabilities of the system, particularly involving people issues
  - All suggestions and feedback were reviewed
  - A business case and future priorities were produced
  - Final specifications were then produced

2006

- System Y went 'live' in February 2006 with basic functions.
- System Y was a document management system (DMS) designed to facilitate the management and administration of corporate documents within Organization BBBBB.
- There were 4 main entities/zones that were used within System Y (see Figure 3):
  - Documentation – all the documents produced were stored.
  - Contact – all contact information held on both internal staff and external clients.
  - Project – all information to projects was stored.
  - Organization – all company information was stored.

2007

Figure 2 illustrates the timeline for System Y's development.

Figure 3. System Y

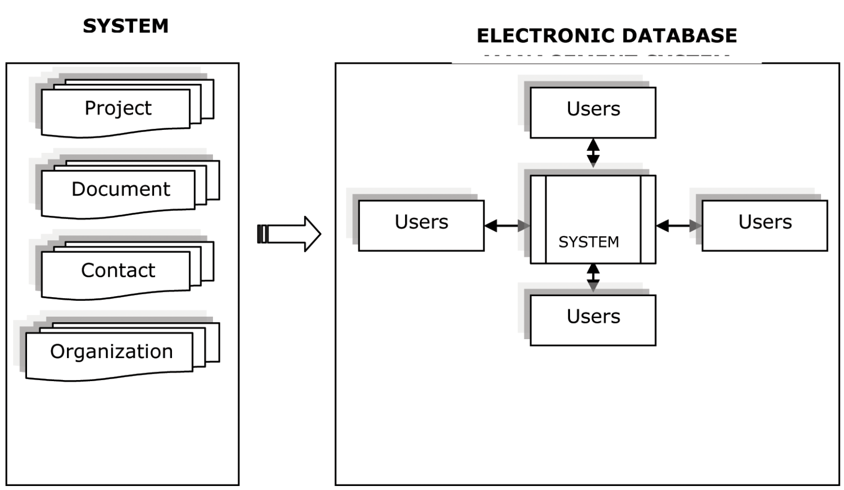
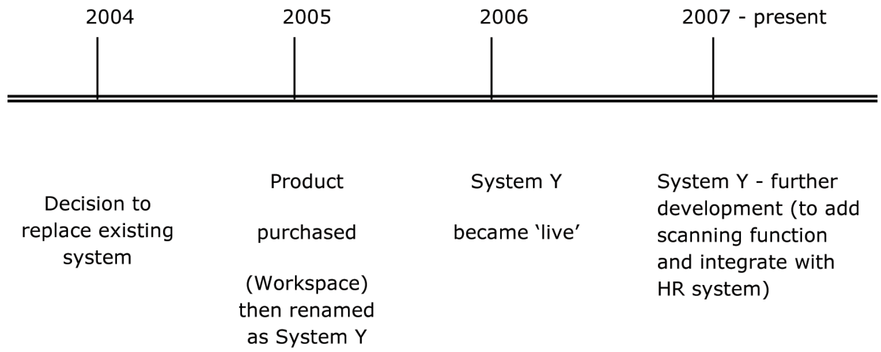


Figure 2. System Y's development timeline



ANALYSIS AND DISCUSSION

The following section addresses each attribute of the four domains of the readiness model within the context of Organization BBBBB.

IT INFRASTRUCTURE

Top Management Perceptions

The management of the company which includes the executive group and senior management group established the vision of strategic thinking towards the general current and future business plan, and the level of resources needed, including staff, training, IT system development, equipment, and identifying opportunities and risks to the business, etc. In addition, any project including construction development and IT system development was planned in terms of its own individual requirements. In terms of IT system development, the business systems manager was responsible for ensuring that IT systems were suitable with business needs and were properly administered and maintained. Once Organization BBBBB became a Limited Liability Partnership (LLP) in April 2006, it facilitated more staff participation in decision making in the company's strategic planning. The *status prior to project* refers to the period prior to 2006 in which the existing Database Management System (DMS) was still in operation. The *current status of the project* refers to when System X became 'live' in 2006 and up to the present day. *Target status* refers to the readiness level that organizations desire to be at in the future.

Status Prior to Project

Drivers

The development of IT/IS is purely to support task requirements and to improve the internal efficiency of various workgroups such as designing, costing, procurement, programming and reporting that is required for the company to be in business. The other IT systems (word processing, spreadsheets, CAD and visualization, computer aided cost estimating and planning and scheduling) are solely independent, off-the-shelf, and primarily related to support individual work tasks. The decision to purchase any system/application was according to the necessity to perform single work tasks. No feasibility study towards the direction of IT/IS development within the company had been done. The need to purchase the system/application was also triggered by looking at systems used elsewhere. A very small IT/IS plan was developed, however its nature was on an ad-hoc basis.

Systems Requirements Definition

The product vendors played an important role in influencing the company to purchase the system/application. The company heavily depended on the vendor to set the functional requirements of the system purchased, while the management determined its non-functional requirements in terms of speed, quality and cost. The applications/systems used within the company were solely dependent on standard featured software and hardware vendors, and operated on a DOS platform. The company did not have any possibility to make alterations or enhancements to the original product they used. The planning, designing and installing of the products was done mostly by vendors.

The Criteria that Relates to the Readiness Model in Level 1

- Driver; copying/work task requirements
- Systems requirements definition; heavily depends on vendor.

Current Status of the Project

Driver

There is an issue relating to the consistency of information storage and retrieval organization-wide. Everyone has their own separate directory and lack of information sharing is obviously emerging. Information queries from the London office take time for a response. The other issue was that all the documents were saved in individual standalone PC, and therefore it took time to get the information needed, and the possibility of lost documents was relatively high. To solve this problem, the management decided to develop a centralized database that was accessible within the company. In doing this, the company recognized the need for a coherent approach to their IT, and the need to bring all the systems under one central control. There was also a space issue regarding the allocation of the filing cabinet for storing the file documents. Cost of reproduction for project documents was also high. The company then decided to develop systems that have access to all parts of the knowledge system through one front end, and be managed by a central database administrator located in Orpington. For example, the applications used Quatopro, Auto CAD, Word, Excel, Visio, and Microsoft Project, and these would be linked together which could then be opened in a single application, 'System Y'. With this feature, the sharing of information and streamlining of communication and inter workgroups co-ordination became more efficient. The main reason to develop System Y was to streamline communication organization-wide, which previous systems did not have the capability to do, and so reduce the costs for document reproduction.

System Requirements Definition

A lot of customization needed to be done to 'Workspace' TM. Organization BBBBB produced a methodology of software development including the necessary adjustment and customization to the software and time frame including renaming the product 'System Y'. The major customization of Workspace TM was to fix the scanning function which facilitates document scanning and viewing. The customization work was done by Vendor B. All the functional requirements and non-functional requirements of System Y were compiled by in-house experts with vendor assistance, and the role of Business System Manager in System Y's development was interpreting the needs of the practice into the technological solution. Vendor B had a project manager dedicated to this practice and information flowed freely between them; there was also a helpdesk function built in. The development of IT/IS started to increase in a planned manner with management support.

The Criteria that Relates to the Readiness Model in Level 3

- Drivers; cost reduction, organization communication
- System requirements definition; mainly in-house with intervention from the vendor.

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

Driver

There is great demand from the clients and project team members to monitor the progress of construction projects more closely and transparently. The future situation was to meet this demand by extending System Y's accessibility to the clients and project team members, and the clients and project team members were given access to System Y to track down the project progress and status. There was also demand from the clients to integrate System Y with their system. Along with these improvements, the agenda to increase internal efficiency was still moving by integrating System Y with the Staff Management System (Human Resource Management System). The future improvement was to add value to System Y by extending its accessibility along the company information supply chain project – this function will increase the organization's competitive advantage.

System Requirements Definition

The company was planning to expand their IT Department by employing more technical staff, particularly the systems analyst, business analyst, and programmer. In the future, the company intended to get full involvement of internal expertise throughout the system's development. The design and set-up of the IT infrastructure such as the operating system, networking, security and telecommunications, was the responsibility of the vendor.

The Criteria Relating to the Readiness Model in Level 5

- Driver; information supply chain project
- System requirements definition; Full in-house with intervention from vendor

The results of top management perception analysis are shown in Table 1.

Systems and Communication

Before 1996, the company had a basic network system but only secretaries had access to it. Quantity Surveyors used stand alone PC's for measurement, etc. There was no internet access or email. Telephones were basic switchboard type and dictaphones were tape type. Photocopiers and faxes were also basic. In 2000, all computers were fully networked and the practice had 80% of staff using computers. Currently, the company has 100% computer access. The telephone, fax and scanner systems are all integrated and managed centrally.

Status Prior to Project

Focus

IT/IS development is to support processes that are required for the company to be in business.

These business operations are primarily related to accounting, designing, procuring, documenting, etc. IT is not critical to any aspect of the company business strategy. The development of IT/IS is purely basic and to support operational tasks and performing similar activities such as typing

Table 1. IT Infrastructure: Top Management Perception Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	1	3	5	2,3,4,5

documentation, payroll, accounts, and human resources. The applications that performed single activities are solely independent, off-the-shelf, and primarily related to the core business such as CAD and visualization, computer aided cost estimating, planning and scheduling.

Network Communication

All the PC's are connected, limited within the group, all employees used their own/shared IT services such as printing and scanning, and all had their own directory to save documents in, that could not be shared by another. However, the documents were exchanged on a manual basis.

The Criteria Relating to the Readiness Model in Level 2

- Focus; co-ordinate project and business information
- Network communication; networks within workgroups in business unit

Current Status of the Project

Focus

IT development is devoted to centralized information activities, company wide. Current goals for IT focus are on cost reduction, improving quality and speed, and enhancing overall company effectiveness. The main aim is to store a document into a single database that is accessible by all staff within the company.

Network Communication

System Y provides a collaborative network for project management through a single login. It authorizes people at all branches/remote locations and the design team involved in the project to access all the information they need, regardless of where the information is actually located. Every PC is connected to each other and shares some common IT services such as printers, scanners and servers, and now the exchange of documents is mostly electronic.

The Criteria Relating to the Readiness Model in Level 3 / 4

- Focus; fully integrated software applications organization-wide
- Network communication; networks are used organization-wide.

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

Focus

IT developments that extend services to support managing the supply chain to the customers and the design team. The IT/IS goals for IT focus on extending the market and the geographic reach. The focus is to make everyone use the same documents such as email, scanning documents, word application, spreadsheets, and drawings to fit around highly organized business processes.

Network Communication

The accessibility to the company database in System Y is extended to the clients and design team.

The Criteria Relating to the Readiness Model in Level 5

- Focus; manage information supply-chain
- Network communication; networks are used to improve supply-chain performance.

The results of system and communication analysis are shown in Table 2.

Table 2. IT Infrastructure: System and Communication Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	2	¾	5	3,4,5

PROCESS

Practices

The overall processes within the organization were not defined and documented prior to the project. Also, prior to the project, everyone followed the same process (scope identified) in managing information; however the processes were unable to ensure quality and standards. Currently the information management has been standardized and everyone follows the same processes, which establishes quality and consistency organization wide.

Status Prior to Project

Practices

The process involved document processing over the whole lifecycle, consisting of incoming documents and 'in house' documents being held in hard copy and kept in dedicated file folders on the filing cabinet with a unique indexing system, documents being scanned when the job finishes, and held on optical disk. When the document was requested, staff could simply search in the filing cabinet according to the unique indexing system and make another copy of it using the photocopier and then the requested document was distributed to the particular staff member. Where documents or data existed on the computer then it could be accessed by anyone with the software on their computer, but searching for it was cumbersome and often required an email to staff members to find out where it was located. The process was repeated every time the same document was requested. There are several issues, which can be highlighted from this process:

- No document to record the document life processing.
- No assigned responsibility and authority for performing the tasks.
- Policies and procedures for managing project information throughout the company were identified and improved based on experience with similar projects.
- The organization typically did not provide a stable environment for managing project information, for example it lacked sound management practices and had ineffective planning.
- The process was unpredictable and constantly changed particularly when dealing with time and cost. The information process capability of the organization was unpredictable, because the management of project information was constantly changed or modified as the work progressed.

The Criteria Relating to the Readiness Model in Level 2

- Practices; business processes scope identified and improved as the work progresses.

Current Status of the Project

Practices

System Y can hold electronic copies of all documents, either in which they were created or in a scanned/pdf form and allows users to create, edit and manipulate documents with the same process and single resource as follows:

- The Find Zones within system Y enable the users to search for and manipulate data. The users are also able to create new projects, edit existing projects, create and edit documents, as well as create and edit contact information.

- The My Zone within System Y gives users access to the latest news, information, and correspondence, active files that users are currently working on, live projects, favorite links, and a list of all documents published, authored, or sent to the current users.
- There are a number of different ways the users can create new documents in System Y; firstly, a new document can be made with the listing templates according to their category, i.e. letters, reports, memos, etc. Secondly new documents can be created via the Project links. Thirdly, by creating a document based on an existing document within System Y; and fourthly, by using an existing document outside System Y as a basis for a new document by dragging a file from Windows Explorer into the Working Files area of System Y.

The process can be summarized as follows:

- The standard process for managing projects information throughout the company are documented and standardized.
- System Y is created with the intention for everyone in the company to be able to store their work and record their activities in a controlled and standard way. For example, every document created such as emails, contact details, word applications, drawings and spreadsheets saved into a single repository can provide a simple and searchable interface to every document. This includes a full content index of most document types. That means the users can search everyone's work to see if something similar has been done before. Also, System Y keeps track of all information relating to the distribution process, thus, it is possible to find out when, how and why documents and drawings have been sent out. Apart from that, global views of all staff and their percentage utilization are also provided to allow the management to monitor the utilization for the entire company. Login and logout functionality exists in System Y, ensuring that two people cannot edit the same document at the same time.

The criteria relating to the readiness model in Level 3

- Practices; business processes documented and standardized

Target Status (Readiness Level that Organizations Desire to be at in the Future)

Practices

The future of System Y's development is to extend its accessibility to the external parties such as clients and project team members to get access into the system. With this function, the clients and project team can track down the progress of the related project, such as variation orders (changes of design, cost increase, extensions of time), payments, purchase orders, valuations, letters, minutes of meetings, etc. This would increase the client and project team member satisfaction and transparency of service being provided by the company, and would increase the level of harmony across the supply chain. Due to its transparency function, a project team is able to make complaints and enquiries in a more efficient and effective manner. A number of customer complaints and enquiries regarding the project progress might be set as a quality measure. However, the external parties' accessibility is limited to the 'project function' area only, which means they cannot view the other functions such as organizational documentation. The future of System Y's development is also intended to integrate with human resource system management for the efficient and effective management of human capital. The human resource management system functions as a single source of information, giving the employees' details such as personal histories, educational background, skills, capabilities, personal experiences, project handling experience, workload to payroll records. This therefore reduces the manual workload of administration activities, such as employee time and attendance, employee tax reports, pension plan to profit sharing and the capability for training and development, skills and capabilities management.

The Criteria Relating to the Readiness Model in Level 5

- Practices; business processes capable of setting quality and measuring the products/services

The results of business process analysis are shown in Table 3.

Table 3. Process: Business Practices Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	2	3	5	3,4,5

PEOPLE

Skills

Most staff gained basic IT skills; however, there are differences between junior and senior staff. The junior staff often request more IT facilities within the company, whereas the senior staff have the opposite view. The company is starting to employ young, dedicated people to fill the IT related positions. Network communication skills are becoming more essential due to great demand from staff and clients to invest in advanced telecommunications.

Status Prior to Project

Type of Skills

Most applications and systems within the company were off-the-shelf and developed by an external consultant/vendor. The in-house IT staff did not have the expertise to develop systems, and in-house staff did the installation of off-the-shelf and low level maintenance. The complex IT maintenance was done by an external consultant/vendor. Most application systems use basic IT functions such as switch on/off computer, creating, saving, and printing documents. The application use is simple and basic and does not require much IT skill. Daily operation of applications such as word perfect and spreadsheets for creating letters, documents, correspondence, specification, project planning, estimating, bills of quantities, etc. Design drawings were done manually. Obviously, no networking and electronic communication skills were required since all the information was distributed through hard copy. The users and even technical staff required assistance from an external consultant in resolving unexpected computer problems.

Capability Building

There was no formal IT training conducted to improve staff IT skills. Users took their own initiative to improve their own IT skills such as learning among themselves and learning through job experience. Some even took external courses on their own. The company only sent technical staff for external training to improve their programming and analytical skills.

The Criteria Relating to the Readiness Model in Level 1

- Type of skills; basic skills
- Capability building; individual effort.

Current Status of the Project

Type of Skills

The original system Workspace ™ (renamed System Y) has multiple functions that were defined by Vendor B. The customization tasks of System Y, such as defining functional and non-functional requirements, were mostly done by in-house expertise along with Vendor B's expertise. In-house expertise was gained with the system development experience during System Y's development. It has to be appreciated that System Y operates for all level of users including senior management, middle management, administration, and site management for Organization BBBBB. Most users, particularly the surveyors, utilize and benefit from the functions within System Y that support many tasks for project management, such as:

- Anyone involved in the project can access the information such as drawings, specification, bills of quantities, letters, etc., with a single source database.
- A project directory stores all the information on parties involved in the project.
- There is a 'template' function for producing all project correspondence including letters, reports, fax's, etc. This ensures the quality of outgoing documents for all projects are maintained.
- Key events for every project such as the start and completion date can be created for monitoring the progress at different milestones.
- Due to the increase in information intensity of the construction process created and used by project members, the company began to realize they could create an environment in which various forms and formats of information could be linked together for easy access and distribution.

Capability Building

The training was conducted throughout the organization including branches, and open to all staff, this consisted of half-day training for all staff to use the basic functions of System Y. Each level of a module is a separate training exercise. When staff successfully complete the basic level of training, they move to the next, more advanced level. Their access to System Y also depends on their training achievement.

The training also conducted for System Y includes

- Series of 18 emails distributed to all staff
- One to one training
    - This is more on an ad-hoc basis and upon request.
    - This is personal one to one training provided to the staff who need more attention about the uses of System Y.
    - The Assistant Training Manager handles this type of session.
- Workshops
    - This is a hands-on workshop about the uses of System Y.
    - The Training/Assistant Manager and Vendor B's trainer conduct this workshop.
    - This workshop is divided into three stages: beginners, intermediate, and advanced.
    - Half day training of all staff to use the basic functions of System Y. Each level of a different module has separate training exercises when the staff successfully complete the basic level of training.
    - All staff were required to pass a lower level training module in order to get to the advanced level. Their access to System Y also depended on their training achievement.

- ◦ Upon completion, the certificates were awarded to the successful trainee.

Training managers give induction courses to new staff, according to the multilevel needs.

The Criteria Relating to the Readiness Model in Level 3 / 4

- Type of skills; considerable technical and project management
- Capability building; centrally integrated training within organization.

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

Type of Skills

Everyone, including the surveyors and administrative staff, expect work under general direction and receive work tasks in the form of project objectives, and are able to plan their own work to meet target costs, time, and quality. Everyone has a good understanding of the relationship between their own area of specialization to other areas. The accounts department is able to understand and appreciate the valuation process.

Capability Building

The existing training methods are expected to be applied for the future and to establish collaboration and knowledge sharing within the company and design team.

The criteria relating to the readiness model in Level 5

- Type of skills; cross disciplinary experience
- Capability building; based on knowledge sharing.

The results of skills analysis are shown in Table 4.

Roles and Responsibility of IT Staff

There are misconceptions about who is really responsible for managing IT within Organization BBBBB between the Business System Manager and the IT Manager. The newly appointed IT Manager was appointed at senior management level in the organization structure. However, the size of the IT Department is small and too inexperienced to cope with a high level IT system.

Status Prior to Project

Position of IT/IS Head

The company had an IT Manager with senior management status.

Roles

IT/IS strategy

The Criteria Relating to the Readiness Model in Level 5

- Position of IT/IS Head; IT Manager with senior management status
- Roles; IT strategy.

Current Status of the Project

Position of IT/IS Head

The company has a dedicated IT Manager, but who is still under the Managing Partner. In addition, the Business System Manager is also in a senior management position.

Table 4. People: Skills Analysis

Status	**Prior to**	**At termination/Current**	**Target**	**Gap**
Maturity Level	1	¾	5	2,3,4,5

Roles

The role of the IT Manager is to deal with all technical matters in IT/IS including the development of the IT/IS infrastructure (hardware, networking and security) and application.

The IT Manager reports directly to the Managing Partner and the staff of the IT Department consists of 7 employees:

- The IT Manager
- 3 technicians
- 1 programmer/analyst
- 2 trainers

The Criteria Relating to the Readiness Model in Level 5

- Position of IT/IS Head; IT Manager at senior management status
- Roles; IT strategy.

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

Position of IT/IS Head

The position of IT Manager is expected to remain the same in the organizational structure.

Roles

The role of the IT Manager is expected to replace the position of Business System Manager in parallel with the expansion of the IT Department and is to help the senior management align the IT/IS strategy with the business strategy.

The Criteria Relating to the Readiness Model in Level 5/6

- Position of IT/IS Head; IT Manager at senior management status
- Roles; IT/IS strategy and business strategy.

The results of roles and responsibility analysis are shown in Table 5.

User Involvement

The awareness of the importance of user involvement is increasing among the senior management. The users did not participate in the decision making to purchase System Y and only participated in System Y's development.

Status Prior to Project

User Involvement

The role of the users is to define the needs and requirements of the proposed system, and only selected users were consulted to give input. No user group was formed to perform tests and the 'champion' of the proposed system normally made all the major decisions.

The Criteria Relating to the Readiness Model in Level 2

- User Involvement; Individual consultation.

Current Status of the Project

User Involvement

A group of users have been selected to be a member in the Steering Group, led by the Business System Manager. Members of the steering committee consisted of over twenty staff members from different levels, with a spread of skills and aptitudes, and from all offices (including the partners to administrative staff) of the practice.

Table 5. People: Roles and Responsibility of IT Staff Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	5	5	5,6	6

The role of the users in the Steering Group is the identification of user needs, including functional and non-functional requirements. The Steering Group had one central meeting then were consulted and involved in specific skills areas up to installation, and the committee exists only for the large IT/IS projects and will be dispersed after the completion of the projects.

The Criteria Relating to the Readiness Model in Level 5

- User Involvement; ad-hoc user group (steering committee).

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

User Involvement

The users are expected to be involved in the systems development by contributing to most of the major decisions, identification of users' needs and requirements, testing the prototype, and approval of the system. However, there is no intention to set up a permanent user group as the practice plan is to go for vendors for full implementation of the future system.

The Criteria that Relates to the Readiness Model in Level 5

- User Involvement; ad-hoc user group (steering committee).

The results of user involvement analysis are shown in Table 6.

Table 6. People: User Involvement Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	2	5	5	3,4,5

WORK ENVIRONMENT

Organizational Behaviour

Prior to the project, the practice saw IT/IS as a tool for performing their work tasks. The IT system and applications purchased developed more focus on specific groups and individuals, and not on the organization as a whole. IT skills and knowledge were measured through the user's competency at using the system, and the relationship between groups was improved in terms of solving the problem. However, there were gaps between the two main groups of users existing in the practice; the 'younger' group who continuously demand more advanced IT systems, compared to the more 'old fashioned' group which preferred the previous practices and systems and were not welcoming towards the new system.

Status Prior to Project

Characteristics

Everyone had their own way of managing information and used a different approach in performing common work tasks such as storing project information in different files and directories. The dissemination of information within the practice took a long time and was mostly done manually which lead to a non-data sharing culture within the practice. The different work processes used in many workgroups within the practice made it even harder. The process also depended solely on individual efforts, such as how long they kept the document until it was filed away, and also how they managed the filing system. Consequently, the company fell into a crisis when a member of staff left the company.

The possibility of document loss was also high. Members of staff without authority were able to take confidential documents without prior consent. The management saw IT as helping to perform business operations, and not as a strategic use; and saw the performance of IT systems in terms of technology delivery (speed, accuracy and quality).

The Criteria Relating to the Readiness Model in Level 2

- Characteristics;
  - Technology led,
  - IT success measured in IT terms rather than impact made on the business.

Current Status of the Project

Characteristics

The practice started to develop a data sharing culture, and there is an improvement of work co-operation within and between workgroups. Everyone has knowledge of what the others are doing. The practice share processes for common tasks such as creating, storing, retrieving and disseminating information inside and outside the organization – such practices have become more transparent. System Y does not change the way people perform their job but just makes the processes easier. IT success is mainly measured in terms of its efficiency, such as how it can make processes run more smoothly. For example, they measured the benefits of System Y by the number of complaints, less time searching for documents, and fewer filing cabinets (operation efficiency).

The Criteria Relating to the Readiness Model in Level 4

- Characteristics;
  - Sharing processes
  - IT is vital for streamlining business processes.

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

Characteristics

The senior management plan to create an IT system environment towards a customer focus and centralized knowledge. The practice sees IT as one of the vital elements that lead to competitive advantage by producing better products and services for the customer. The practice also encourages more user participation in IT system development in the future and users are expected to become a main reference during IT system development projects. The interaction and relationships between the users and IT staff is improving in parallel with the plan for IT department expansion in the future.

The Criteria Relating to the Readiness Model in Level 5

- Characteristics;
  - IT/IS used to provide better products and services to customers
  - Staff encouraged to input ideas.

The results of organization behaviour analysis are shown in Table 7.

IT Policy

Organization BBBBB centralized all of their IT/IS activities including purchasing and managing the IT infrastructure and applications. Previously, prior to the establishment of the IT department,

Table 7. Work Environment: Organization Behaviour Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	2	4	5	3,4,5

the quantity surveyor (currently appointed as Business System Manager) controlled the management of IT across the organization. Currently, the IT department controls the IT/IS activities across the organization. However, the IT Manager has to report to the Business System Manager on every decision they make. The Business System Manager is the 'champion' of the IT/IS management across the organization and the management is expected to practice the current policy on managing their IT/IS activities in the future.

Status Prior to Project

IT/IS Activities Control

Mostly controlled by Business System Manager.

The Criteria Relates to the Readiness Model in Level 3

- Centralized IT infrastructure and application policy.

Current Status of the Project

IT/IS Activities Control

The IT department controls all IT/IS activities (centralized IT infrastructure and applications). Every time the business units want particular applications they have to submit applications to the IT Manager, making a business case. If it is not considered expensive, then the IT Manager will discuss with the group leader and then get approval and install. Otherwise, the IT Manager will require approval from the Managing Partner.

The Criteria Relating to the Readiness Model in Level 3

- Centralized IT infrastructure and application policy.

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

IT/IS Activities Control

The IT department controls all IT/IS activities (centralized IT infrastructure and applications).

The Criteria Relating to the Readiness Model in Level 3

- Centralized IT infrastructure and application policy

The results of IT policy analysis are shown in Table 8.

IT Policy

There was a lack of senior management awareness towards the potential of IT for their organization. Previously, the IT development was organized on an ad-hoc basis. System Y implementation was the first time the IT development in the practice had gone through the proper planning process. The senior management sees IT/IS as tools for smoothing the business processes, and the Business System Manager plays a vital role in representing senior management's views.

Status Prior to Project

Communication

The communication regarding the IT/IS activities was mostly done through manual methods such as through memos, letters and meetings. Most communication between senior management and

Table 8. Work Environment: IT Policy Analysis

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	3	3	3	Improvement

staff was through the Business System Manager, and the communication between the vendor and the users was also through the Business System Manager. No communication planning existed and everything was done on an ad-hoc basis.

Participation

There was little participation of senior management in IT/IS activities and the senior management were mainly concerned with the IT/IS expenditure.

The Criteria Relating to the Readiness Model in Level 2

- Communication; no established standard for communication (ad-hoc)
- Participation; ad-hoc participation.

Current Status of the Project

Communication

The communication regarding the IT/IS activities including IT/S project is mainly done through emails and during training. Apart from that, communications through memos, letters and meetings is still popular. The senior management began to communicate to the users directly through emails and meetings. The senior management comment on, and answer, any question from the users through email. E-mail played a major role as a communication channel during System Y's development. For example, the introduction and guidelines on using System Y were done through a series of 18 emails. The vendor also conducted an interview with the users to seek out their needs and expectations towards System Y. However, only selected users were chosen to participate in the interview. The practice started to have a communication plan for IT/IS development.

Participation

In addition to the expenditure on IT/IS, the senior management were concerned with the disruption caused during System Y implementation to the business process. The senior management also started to express their concern regarding the improvement/enhancement of a future System Y to improve client satisfaction, while the Business System Manager participates actively in System Y development.

The Criteria Relating to the Readiness Model in Level 3

- Communication; organization-wide policy and standards for communication
- Participation; Participate only on large scale and high cost IT projects

Target Status (the Readiness Level that Organizations Desire to be at in the Future)

Communication

The practice expected to develop communication planning across the organization for their IT/IS activities, from the experience gained in System Y's development.

Participation

For System Y to successfully satisfy the clients, the senior management have to play a vital role to participate in measuring the performance of the IT/IS project from the strategic perspective.

The Criteria Relating to the Readiness Model in Level 4/5

- Communication; a communication plan is expected for all activities
- Participation; Participation in performance measurement of IT/IS project.

The results of Organization BBBBB behaviour analysis are shown in Table 9.

Table 9. Work Environment: Leadership

Status	Prior to	At termination/Current	Target	Gap
Maturity Level	2	3	4/5	3,4,5

SUMMARY AND FINDINGS

Overall processes for managing information are improving (level 2 to level 5 - process) since the introduction of System Y where;

- Overall processes were not defined and documented prior to the project.
- Also, prior to the project, everyone followed the same process (scope identified) in managing information; however the processes were unable to ensure quality and standards.
- At the current status, the information management began to be standardized and everyone followed the same process which established quality and consistency organization-wide.

Prior to the project, the development of IT/IS was driven to perform daily work tasks that required the company to run a business. Examples of those work tasks are designing, costing, procurement, programming and reporting (level 1 – top management perception). Organization BBBBB then developed System Y to streamline communication organization-wide, which the previous system did not have the capability to do, and to reduce cost for document reproduction (level 2, 3 – top management perception). System Y development is currently continuing to achieve its target for managing the information supply chain within industries (level 5 – top management perception).

The level of IT skills prior to the project was relatively low (level 1 - skills). This was due to the fact that most of the application systems used basic IT functions such as switching the computer on/off, and creating, saving and printing documents. The application use was simple and basic and did not require much IT skill. The daily operations were of applications such as word perfect and spreadsheets for creating letters, documents, correspondence, specifications, project planning, estimating, bills of quantities, etc. The introduction of System Y has helped the company to increase their staff's IT skills. For example, the customization tasks of System Y such as defining functional and non-functional requirements are mostly done by in-house expertise, along with Vendor B's expertise. The in-house IT/IS experts also gained system development experience during System Y's development. The staff's IT skills are expected to improve in the area of using IT for decision making, and across disciplinary skills, parallel with the further enhancement of System Y in the future (level 2,3,4,5 -skills).

Prior to the project, the position of the IT Manager was at senior management level to set up IT/IS strategy. The Business System Manager was also at the senior management position (level 5 – roles and responsibility of IT staff). The position of the IT Manager remained the same during System Y's implementation and then remained so, but the role of the IT department is now to set up IT/IS and business strategy (level 5, 6 - roles and responsibility of IT staff).

There was no improvement in IT/IS activities control since it remained centralized, for all of their IT infrastructure and applications(level 3 – IT policy). In the improved system, the users get involved in the IT/IS development, which is an improvement from the ad-hoc individual participation of the previous system (level 2 – user involvement). They get participation in the steering group led by the Business System Manager in System Y's project (level 5 – user involvement). However, the Steering Groups only exist for the large IT/IS projects, and will be dispersed after the completion of the projects.

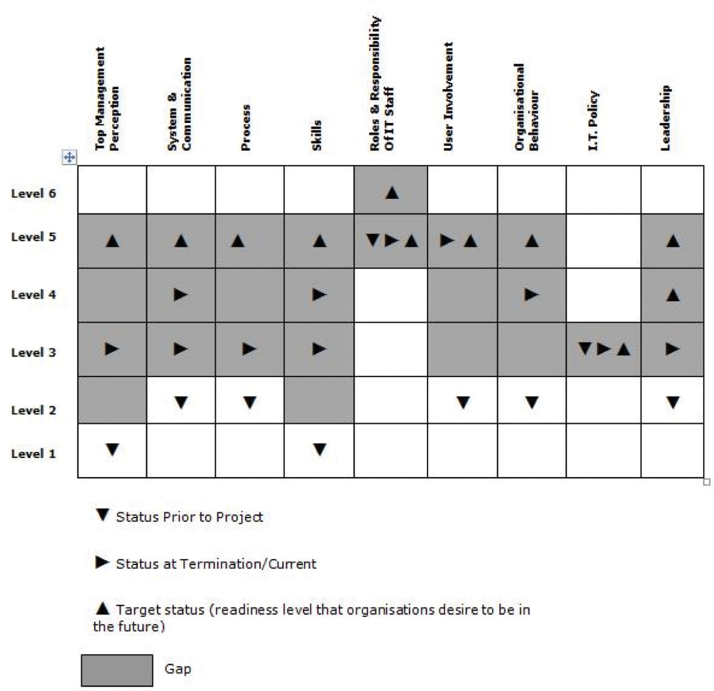
Two groups exist within the company. There is a gap between the two main groups of users existing in the practice. They are the 'younger' group which continuously demand a more advanced IT system, compared to the more 'old fashioned' group which prefer the previous or current practice. Prior to System Y's implementation, the management saw the performance of IT systems in terms of technology delivery, i.e. speed, accuracy and quality (level 2 – Organization BBBBB behaviour). However, this perception changed after System Y became 'live'. The management sees IT/IS as vital for smoothing the business process and as a tool for achieving competitive advantage (level 4,5 – Organization BBBBB behaviour).

The summary of the assessment of System Y's implementation according to the readiness model can be found in Figure 4.

CASE STUDY ADOPTED FROM

Salleh, H. (2007), "Measuring Organisational Readiness Prior to IT/IS investments", PhD Thesis, School of the Built Environment, The University of Salford,UK

Figure 4. Summary of the assessment of System Y's implementation



This work was previously published in Cases on E-Readiness and Information Systems Management in Organizations: Tools for Maximizing Strategic Alignment, edited by Mustafa Alshawi and Mohammed Arif, pp. 84-113, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 100
Embedded RFID Solutions: Challenges for Product Design and Development

Álvaro M. Sampaio
Polytechnic Institute of Cávado and Ave, Portugal & University of Minho, Portugal

António J. Pontes
University of Minho, Portugal

Ricardo Simões
Polytechnic Institute of Cávado and Ave, Portugal & University of Minho, Portugal

ABSTRACT

Full traceability of products is extremely difficult, although it has been sought after for as long as production, distribution and sales chains exist. Electronic traceability methods, such as RFID technology, have been proposed as a possible solution to this problem. In the specific case of RFID, the number of applications that promote innovative solutions in retail and other areas has been continuous growing. However, RFID tags are mostly placed externally on a surface of products or their packages. This is appropriate for logistics, but not for other applications, such as those involving user interaction. In those, not only is the placement of the RFID tag more complex, but it is also necessary that the tag is not visible or not directly accessible, to prevent accidental damage and intentional abuse. This certainly imposes challenges to manufacturing, but mainly creates new challenges to the development of new products and re-design of existing ones. This chapter presents some insights and what we consider to be the two main approaches to incorporating RFID technology into consumer products.

INTRODUCTION

RFID technology has attracted an increasing interest from companies and R&D institutions around the world in the past few years. Some of the application areas of this technology can be seen in the automotive, retail, logistic and health industries. The benefits can go from increased productivity and cost reduction, to more indirect factors– difficult to quantify – such as improved post-sales consumer service (Hodges & McFarlane, 2005). This technology is seen as a potential new phase

DOI: 10.4018/978-1-4666-1945-6.ch100

in the development of the information society, in which the overall day-to-day objects will be inter-connected to other objects, data, etc. This will lead to a considerable impact in some value chains, resulting in a plethora of new applications (Gellersen, Schmidt & Beigl, 2000).

However, this trend represents a challenge that has limited the feasibility of applying RFID technology. The fragility of these electronic devices, associated to the great variety of products and the processes to which these are submitted through their life cycle, restrict the adoption of these technologies (Deavours, Ramakrishnan & Syed, 2005), Moreover, the development of solutions to incorporate an RFID tag in a product and/or packaging in molded components with the minimum impact on the manufacturing process still remains a difficult task (Hodges & McFarlane, 2005).

Despite recent evolutions in design and development processes, that cover the necessary modular product architecture (e.g., Grady, 1999; Ulrich & Eppinger, 2007), these present themselves as processes for new or re-designed products. None of these new processes have been optimized for the development of products that need to perform their function coupled with another product that already exists, being therefore conditioned by that existing product's form, size, and other physical and functional characteristics.

Thus, the challenges posed by introduction of RFID technology into consumer products can be divided into two types: those related to the manufacturing process (the ability to embed the RIFD tag into the product), but also those related to the required changes to the product form and possibly its architecture (where in the product should the RIFD tag be placed and how much does the product need to change to accommodate it). For the former, we can expect industrial processes to improve to meet the demands of the market. As RFID technology becomes more widely adopted, advanced processing technologies (such as non-conventional injection molding) will enable new solutions for embedding the tags into products during the manufacturing process. However, for the later, which is the focus of this work, it will clearly be necessary that appropriate design methodologies be identified and, if necessary, developed, for seamlessly adding RFID tags to products.

In this chapter, we discuss that to successfully incorporate RFID technology, the design and development of products will demand an effective analysis of existing products, following a modular product architecture philosophy, which will create a significantly different development process. We also discuss what those changes may imply in the overall process and which type of product architectures offers competitive advantages for such solutions.

RFID TECHNOLOGY OVERVIEW

RFID is a wireless tracking technology that allows a reader to activate a transponder on a radio frequency tag attached to, or embedded in, an item, allowing the reader to remotely read and/or write data to the RFID tag (Das, 2009). Essentially, an RFID system comprises an RFID tag, also called transponder, a reader, also called transceiver, and the supporting IT infrastructure. An RFID tag embodies a built-in antenna connected to an electronic microchip. These tags carry on a unique identifier that relates the tag with the precise tagged object, allowing unique product identification. A tag receives and retransmits signals on a set of predetermined frequencies, in other words, in response, a tag transmits a predetermined message to a predefined received signal.

RFID tags can be read-only or read-write. A read-only tag includes a programmed identification code, recorded at the time of manufacture or when the tag is allocated to an object. Once programmed, the data cannot be modified but may be read multiple times. Read-write tags can have their memory changed, or written many times; in

Table 1. A general overview of RFID tags

	Passive	Semi-Passive	Active
Power Supply	External	Internal battery powering only processing units and sensors	Internal battery also powering radio chip
Typical read Range Requirement	1-3 meters	Up to 30 meters	Up to 120-150 meters
Typical Type of memory	Read-only	Read-write	Read-write

this way, they offer greater functionality. RFID tags can be categorized in three general varieties: passive, semi-passive and active tags (Table 1).

A passive RFID does not contain an internal power supply. This tag transforms the energy of radio-frequency it acquires from an RFID reader to answer by sending back information. Drawing power from the reader makes them cheaper and smaller than active tags, which have a battery used to broadcast the signal to the reader (Teixeira & Piçarra, 2007). Active tags are also more reliable than passive tags, because they have an internal read and write capability, and can transmit their signals over a longer distance. Depending upon the size of the tag and the frequency used, the current range of reception, or read range, of the reader is limited. Semi-passive tags use an internal battery to ensure data integrity, however the signal sent from the reader generates the power to transmit the signal from the tag.

The fast, automatic, pervasive and ubiquitous identification of objects is one of the challenges of today's business corporations. RFID's ability to read objects in motion and without a direct line of sight gives it the edge over traditional bar-coding methods. It can be said that the difference between RFID technology and bar-coding comes with the sensing capabilities that RFID offers in an increasingly dynamic, mobile physical world (Teixeira & Piçarra, 2007). Even if the main differences between these technologies may be seen as minor, other differences are opening the path for new applications. If we imagine that any product could have an RFID tag and that we could track this product with a computer-reader, this can be translated in an incredible number of applications in various domains, such as access control, logistics, or animal tracking. Some of the most interesting applications of this technology are: (a) the counterfeiting battle, which is huge threat to global business and concerns all kind of products and companies (e.g. pharmaceutical, luxury goods), as RFID technology allows the products to authenticate themselves to the user; (b) the reducing of crime, under the Design Against Crime program (Learmount, Press & Cooper, 2000; Design Council, 2002), which has been producing some interesting examples since 1999, such as the *Bike-off* project (Bikeoff, 2009) and *The Chipping of goods Initiative*, to show how property crime can be reduced through RFID technology (Adams & Hartley, 2000); (c) e-Government, as there is a growing demand for higher security standards and governments requiring secure identification of individuals, such as electronic passports; (d) e-Health, with healthcare being one of the most enthusiastic sectors adopting this technology as it offers multiple possibilities of tracking medical devices and; (e) Waste management. It is precisely this last example that emphasizes another important factor in the development and implementation of RFID technologies. Nowadays, as sustainable development is increasing in importance and every company must worry about their products throughout the end of the product lifecycle. It is very difficult to create a disassembly system for any kind of a product. It would be better if every company could trace its products and find out where they are at the end of their lifecycle. As stated, while being an innovative solution itself, it can be said that RFID triggered further innovation in products, services as well as in development processes.

RFID IN CONSUMER PRODUCTS

Incorporating RFID tags in products might seem like a trivial task. To support that hypothesis, one can even showcase the large number of products which already include RFID tags. However, a careful analysis shows that this is not the case. First, the large majority of current RFID-tagged products are containers (e.g. boxes, palettes) with very simple geometries. Second, in almost every case of currently tagged products, it is actually the product package which is tagged, and although packaging is a major application of RFID, and many products have and will continue having a package, it is important to be able to place tags in products that are commercialized without a package. Due to these two factors, until now there has not been a serious need to develop and implement efficient methods to place RFID in a wide variety of products. With the massive increase in RFID deployment and expansion to new applications/ products, this situation must change very soon.

The two main reasons why it is important to start studying in much greater detail the placement of RFID tags in products is due to: a) technological issues related to the RFID technology itself, and b) the variety in products to be tagged. Other very important aspects of RFID incorporation into products are cost and sustainability, but these are outside the scope of this work and will not be explored.

The main RFID technological issues/limitations are signal attenuation and electromagnetic interference. Signal attenuation in RFID is related to losses in the energy reflected by the tag back to the reader which sends the signal. If less energy is able to reach the tag, then the tag must be closer to the reader to be read. The reflected signal decreases as the inverse fourth power of the distance between tag and reader. Since most tags are passive, the energy levels are already small to start with, and if losses occur, read distances can become a problem. Attenuation can occur from the environment or product characteristics. Water (thus, also products with high water content, including fruits, soft drinks, or moist goods) and carbon will absorb ultra-high frequency RFID signals (the most common passive RFID tags). In the case of metals (canned goods and metal objects), the RFID signal will bounce off the metal surface, typically causing the signal to be blocked. EMI (electromagnetic interference) is essentially noise that will affect the RFID signal travelling back to the reader. This can result from interference from motors, conveyors, old wireless networks, and other sources. In most cases, the source of the interference needs to be shielded in order to allow RFID system to work properly.

Signal attenuation requires particular attention due to the large variety of materials currently used, both in the product itself as well as its package. The development of consumer goods has evolved to take advantage of materials engineering, and thus contemporary products use materials such as wood, plastics, metal, paper/cardboard, composites, laminates, glass/ceramics, and often include several parts made of different materials. Without understanding the influence of the materials on RFID performance, it is not possible to ensure adequate implementation of this technology.

The second reason listed above for studying RFID placement in greater detail was the large variety of existing products. The diversity of materials has already been discussed, but product shape and other characteristics are also an issue. For example, a material might be relatively impervious to the RFID signal but only up to a certain wall thickness. Regarding shape, qualitative and quantitative criteria must be defined for where to place the tag on a product, and where along the production line the tag should be incorporated (possibly, but not highly likely, at the end). Trends towards more complex or more organic product shapes will affect where a tag can be placed, and specific inner features might be required to accommodate a tag. This becomes even more important due to the fact that specific design solutions can solve some of the signal attenuation issues listed above. For example, although metals can create serious problems for direct tag placement, introducing an air gap between the tag and the metal surface can increase its performance, in some cases, bringing it close to the level of performance expected in materials to which the RFID signal is impervious.

Other issues can also affect the RFID performance, for example the proximity of neighboring tags and the angle between the reader antenna and the tag. If RFID is being used for automated inventory upon delivery of goods to a store, adequate planning of the position of the tag in the product and the way multiple products stack inside a box can optimize the performance of the system.

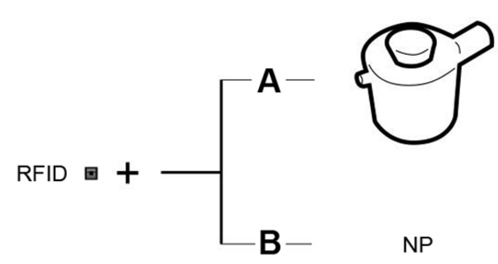
It is also interesting to notice that different sectors of activity have been employing RFID technology in very different degrees. In retail, one can assume this to be highly related to considerable differences in the type of product of each sector. For example, textiles, consumer electronics, shoes, food packaging, all require considerably different implementation solutions, and a common platform is difficult to achieve. Due to this, the massive deployment of RFID in consumer products that has been advocated in the past decade is taking much longer than expected to come to pass. This situation results mostly from the inexistence of adequate methods to plan RFID tag placement in products, prevent knowledge-based development.

How to incorporate RFID tags in products in the most efficient way is a problem that needs to be solved. Two main approaches can be used for this. Externally attach the RFID tag to the product in some way, or make the RFID tag part of the product itself.

THE PRODUCT DEVELOPMENT PROCESS: TOWARDS THE INCLUSION OF RFID TECHNOLOGY

In the general trend towards wider application of RFID technology in products, we can envision two different roles that product development will assume. If we can think that tagging will move from being something attached to an object to become part of the object itself; for this to happen we will inevitable see changes take place in the design and product development process. Nevertheless, tagging any product is not a simple task and, thus, the shift from tagged to RFID-embedded products will be a necessary evolution. As stated before, product design and development processes present them as process for new or re-designed products. However, none of the traditional processes meet the requirements of developing products that support the incorporation of an RFID tag. During the product development process that absolutely requires the inclusion of RFID technology, we need to develop secondary products that have an embedded RFID and can be coupled to existing products. For this we need two product development processes: one that can add RFID technology to products that already exist (product development process – A), and a different process that has the incorporation of RFID technology in the product as one of its primary concerns (Product development process - B). These are schematically represented in Figure 1.

Figure 1. The two possible RFID product development process. NP stands for "New Product"



Defining the development process as the specific series of events and methods by which a procedure or set of procedures are followed, in order to achieve an intended purpose, goal or outcome (Best, 2006), the development process consist in a series of activities and methods that are grouped in phases that tackle a complex problem by decomposing the overall complexity in smaller, more manageable sub-problems (Simon, 1969). Dividing a problem into simpler sub problems is called decomposition and there are many schemes by which a problem can be decomposed. One of them is functional decomposition (Ulrich & Eppinger, 2007). In product development, a part of the problem-solving activities consists of discovering unknown types of interdependencies between different design solutions. A decomposition of the product into nearly independent functional components is a way to clarify the overall problem and test the feasibility of different design solutions. A product can be thought in both functional and physical terms. The functional elements of a product are the individual operations and transformations that contribute to the overall performance of the product and the physical elements of a product are the parts, components, and subassemblies that implement the product functions (Ulrich & Eppinger, 2007).

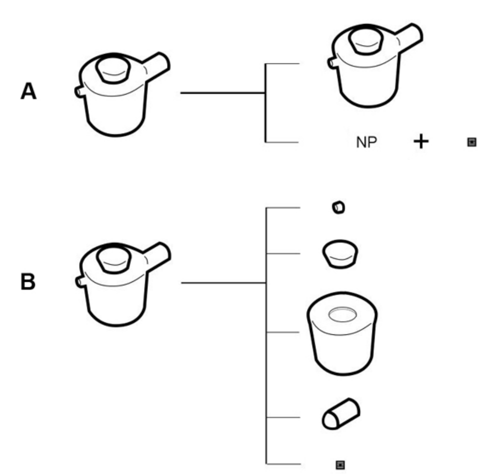
Allocating a function of a product to a physical component is called product architecture. Roughly formulated, product architecture concerns the design and layout of the physical building blocks of a product. Product architecture is an abstract way of looking at a product and is usually determined before a product is actually designed into detail (Ulrich & Eppinger, 2007). Likewise, abstract terminology is used to refer to properties of components in products and the way these components interact. Ulrich and Eppinger (Ulrich & Eppinger, 2007) specified product architecture more precisely as: (a) the arrangement of functional elements; (b) the mapping from functional elements to physical components, and (c) the specification of the interfaces among interacting physical components.

The importance of problem decomposition and the subsequent product architecture to the inclusion of RFID technology is very high. Once again, if we need to place an RFID in existing product and further start to embed this technology in new products, we need to understand the existing products, their shapes, forms and functions to design a way to incorporate the RFID tag. Rethinking the product architecture is the best strategy to attach a new physical component, with a specific function, to an existing product.

In other words, when we need to attach an RFID tag to an existing product, and we begin with functional product decomposition, the existing product becomes one functional element itself. What is new in that architecture is the RFID tag. To attach the RFID tag to the product we need to develop a new complementary product (NP) which contains the RFID tag (process A in Figure 2). This new complementary product will attach itself to the existing product, which ensuring adequate physical support to the RFID tag.

On the other hand, process B in Figure 2, illustrates the functional product decomposition of a product which is being developed and in which we need to incorporate the RFID technology. The

Figure 2. Functional product decomposition



fact that we have a considerable degree of control over each of the physical elements, with the exception of the RFID tag, will allow us to embed the RFID tag directly in the product. Independently of the product development process employed, one of the physical components will have the embedded RFID tag, ensuring traceability for the entire product. This, of course, will lead us to modular product architecture.

THE MODULAR APPROACH

According to a widely accepted definition (Ulrich & Eppinger, 2007), a product architecture is modular when components are functionally independent. A modular architecture includes a one-to-one mapping from functional elements in the function structure to the physical components of the product, and specifies de-coupled interfaces between components (Ulrich & Eppinger, 2007). Functional independence has a deep impact on the supply chain, since components may efficiently be developed and manufactured separately by different organizations, as well as on marketing, since product customization may be achieved with greater ease by simply swapping components (Cantamessa, 2005) One of other characteristics of product modularity is the ability to mix and match independent and interchangeable product building blocks with standardized interfaces and products in order to create product variants.

Modularity is an effective means to control complexity. Once the modular architecture is fully specified, the uncertainty in product development is confined to problems of identifying feasible solutions that will improve the functionality of the product, allowing at the same time the firm to minimize the physical changes required to achieve a functional change (Ulrich & Eppinger, 2007). Furthermore, it enables the production of variety while facilitating the achievement of economies of scale. Modularity is vital to responsive manufacturing and creates the opportunity of producing a huge variety of products that satisfy various customer requirements. A firm that offers several variants of a product can, at the same time, postpone the differentiation of a product until the very end of the supply chain. This is because products manufactured with a modular architecture can be easily modified without causing significant changes to the manufacturing system, because of the possibility of sharing components amongst various products.

The provision of product variety is often based on the concept of product platform. Since the effort, in terms of time and money, to implement a product platform is larger than a traditional development process, it is not possible for any new project to necessarily be a platform project. Product platforms are closely related to the technology development efforts of a determined firm, and to decisions about which technologies to employ in new products (Cantamessa, 2005). The technologies and the variable products needed, or possible, are expected to be planned in the beginning of the development process. This is the case of the inclusion of RFID technology in products. With the product platform being the set of assets shared across a family of products (Robertson & Ulrich, 1998), RFID technology is one of the assets shared, and needs to be planned in a way that achieves the overall performance of the technology. In other words, RFID technology could be the base for a new product platform, independently of the strategic development process chosen.

Notice that in the framework of the two distinct product development processes discussed above, the practical implementation of the integration of RFID technology in products can range anywhere from process A to process B. The process which can be implemented more rapidly and more straightforwardly is Process A, since it does not imply changes to the product itself. However, on a longer time scale, Process B will most often be employed, due to its inherit advantages.

If we assume that RFID technology will be the support for one or several product platforms, the planning of these platforms will necessarily involve managing a basic trade-off between distinctiveness and commonality. These trade-offs can be formalized in the differentiation plan and the commonality plan (Ulrich & Eppinger, 2007). The differentiation plan explicitly represents the ways in which multiple versions of a product will be different from the perspective of the market or the customer. These differences are normally diverse specifications of the final product with the objective of satisfying the needs of the customers in a specific market target. Conversely, the commonality plan explicitly represents the ways in which the different versions of the product are the same physically and, sometimes, functionally. It is clear that in both product development processes discussed in this chapter, establishing the architecture of the product and its inherent modularity is a primary concern in the overall process. Nevertheless, turning to a more specific modular philosophy will necessary inevitably lead us to planning a product platform around RFID technology. In the case of the product development process A, it is necessary to take into account that developing a complementary NP that only works for one existing product is (aside from exceptions) unreasonable in terms of an economy of scale. So the inclusion of an RFID tag through this product development process will require a profound analysis of the shape of a variety of products (probably in the firm internal product catalogue) with the purpose of finding similarities that make it possible to develop a unique NP with an embedded RFID that can attach itself to as many products as possible (Figure 3).

From Figure 3, one can see that one of the possible similarities in the four products illustrated is the handle. Thus, to insure the possibility of incorporating RFID technology, it is necessary to create the NP with the embedded RFID in such a shape as to allow it to attach to all the handles of the considered products. This could be done with, as an example, a flexible polymeric bracelet.

On the other hand, product development process B (Figure 4), considering the development of new products from start, it is not restrained to any shape, form or style, but only to the functional elements needed. As in process A, several possibilities can be identified to achieve the goal of developing a new product that already features an embedded RFID. After decomposition of the functional elements, one needs to identify which functional element is going to have the embedded RFID; in the case shown in figure 4, that element was the cap.

Thus, in this case it is possible to talk about a functional element (asset) common to all products, and which already has an embedded RFID tag. It can also be planned to perform its function in other product families.

Figure 3. Possibilities of product development in the scope of Process A

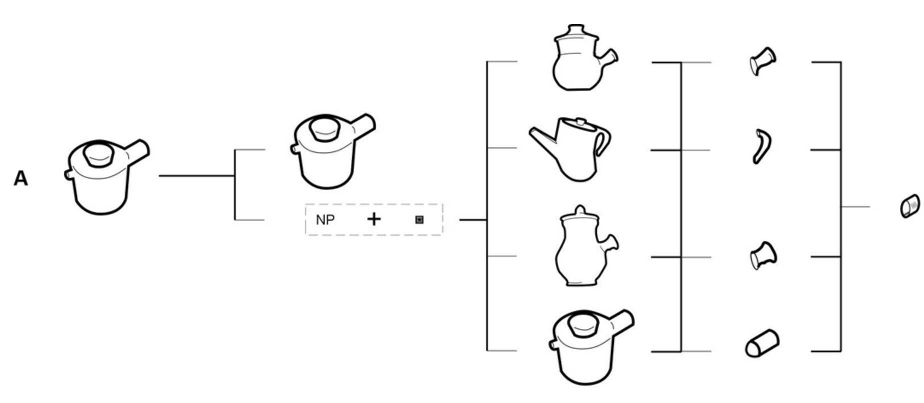
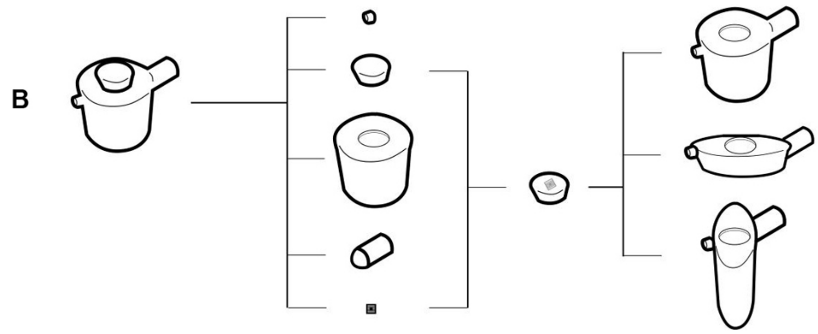


Figure 4. Possibilities of product development in the scope of Process B



CASE-STUDIES

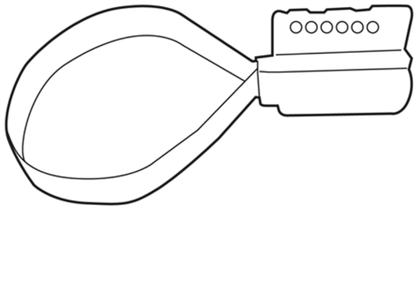
To better illustrate the two concepts being the product development processes discussed in this chapter, we shall use two case-studies of commercial products. The two products are mass-produced by the Daily RFID Co, Ltd. (2009).

The first example (shown in Figure 5) is a bracelet-like product that was designed to attach RFID tags to existing products, following the approach of product development process A.

The shape of the existing product to which this device will be attached must include some feature around which the bracelet will go (e.g. a tubular or quasi-cylindrical feature). It is a one-off seal, and once attached, cannot be removed without damaging the seal and leaving evidence of tampering.

As is well known, the most pervasive current application of RFID technology, and the one that still represents the large majority of tags in use, is logistics. Essentially, RFID tags are placed in boxes, cases, pallets, carts, and more recently, envelopes, books, magazines, and many other objects which have very simple shapes, and, almost without exception, offer a flat surface for placing the tag. For all those applications, the typical tag offers an adhesive layer so it can be stuck to the surface. The type of device in Figure 5 (which is only representative of a large number of devices with encapsulated RFID tags), appears from the need of placing tags in more complex products. If a retail store wishes to track its products that are not enclosed in a box (e.g. bicycles, toys, kitchen accessories, etc), they must find a way to ensure the tag can be safely placed on the product and that it will not come off when colliding with other objects. When there are no flat surfaces or when it is not allowed to leave residues of the adhesive on the object, the type of device in Figure 5 can be an excellent solution. It also prevents malicious users to simply peel off a tag placed externally on a surface of the product.

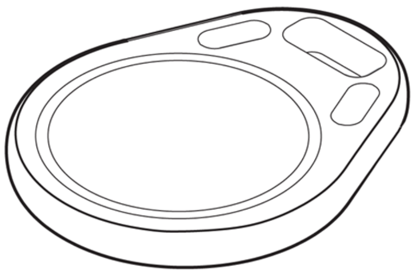
Figure 5. RFID embedding solution for external coupling to existing products



The second example is a key-chain (Figure 6). Designed from the start to include an RFID tag (following the approach of product development process B), using adequate materials and some planar surfaces, it becomes possible to use the key-chains for access control, security, and many other applications. The key-chain can even allow information (e.g. a logo) to be printed on its surface.

In this case, the fact that the RFID tag is not detectable by users is a vital advantage. At the

Figure 6. Key-chain with embedded RFID



early design stages, aside from all the product design specifications that one would typically consider for such product, the shape of the key-chain must be set taking into consideration physical requirements of the RFID tag as well as its placement. This implies finding the optimum way to enclose the tag inside the key-chain.

There are many technological options (fabrication solutions) to enclose the tag inside the key-chain, such as making the keychain in two parts that are locked together only after placing the tag between them in an appropriate hollow space, or finding a way to place the tag actually inside the polymeric material during processing. Notice that, in practical terms, the mechanism through which the tag is embedded in the product is not critical, since the end result is essentially indistinguishable by users.

CONCLUDING REMARKS

RFID technology is at a crucial point, in terms of the adoption exceeding current applications. As this technology becomes ever more deeply embedded in everyday life, it can no longer be seen as a simple appendix to increase traceability in the supply-chain, but as a technology that can be embedded in every existing product. The technical issues of tag reliability and the effect of liquids and metal surfaces are challenges and not true problems. One of the most important issues in the implementation of this technology is that the RFID will necessary transform the way in which we develop products and the way we identify potential advanced features of products. One can say that very little RFID novelty will be achieved from changes in the technology itself. Instead, it will arise from innovative use. When it comes to finding new ways to use RFID to create and build product value, user experimentation will inevitably bring original interactions.

In a trend towards a more effective use of this technology in product design and development, RFID tag incorporation needs to be an important specification of the development process. This, of course, will be an evolutionary process, and will range from the previously described development process A, where no changes are made to existing products, through development process B, where products are developed from start considering embedding the RFID tag. The endgame in any of those processes will be the increasing use of RFID technology in products, increasing the interaction of products with their environment and their users. A time will come when RFID will vanish from the mind of the consumers, becoming impossible to distinguish it from the product itself. As espoused by Weiser (Weiser, 1991, p. 94), "the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it".

Although the discussion about the examples provided in previous sections illustrates the need for manufacturing processes optimized for embedding RFID technology, processing technologies exist that can be adapted for that purpose, and the real challenges still clearly reside on appropriate product design and development methodologies.

It is also important to mention that RFID is only one example of the growing number of microelectronic devices incorporated into products, following the current trend of technology-driven product development (Simoes & Sampaio, 2008). However, many of the challenges are the same, and conceptual solutions can be reused for different scenarios.

ACKNOWLEDGMENT

We acknowledge the Foundation for Science and Technology, Lisbon, through the 3° Quadro Comunitario de Apoio, and the POCTI and FEDER programs.

REFERENCES

Adams, C., & Hartley, R. (2000). *The Chipping of goods initiative. Property crime reduction through the use of electronic tagging systems. (Police Scientific Development Branch, Publication Nº 23/00)*. London: Home Office.

Best, K. (2006). *Design Management: Managing Design Strategy, Process and Implementation*. Lausanne: AVA Publishing.

Bikeoff. (n.d.). *Bikeoff*. Retrieved April 9, 2009, from http://www.bikeoff.org

Cantanessa, M. (2005). Product portfolio management. In Clarkson, J., & Eckert, C. (Eds.), *Design Process Improvement – A review of current practice* (pp. 404–435). London: Springer.

Daily RFID Co, Ltd. (n.d.). *Daily RFID*. Retrieved April 9, 2009, from http://www.rfid-in-china.com

Das, R. (2002). *An Introduction to RFID and Tagging Technologies*, Cambridge, UK: White Paper, IDTechEX

Deavours, D., Ramakrishnan, K. M., & Syed, A. (2005). *RFID Performance Tag Analysis (Tech. Rep.)*. University of Kansas, Information and Telecommunication Technology Center.

Design Council. (2002). *Evidence Pack. DAC case studies*. London: Design Council.

Gellersen, H. W., Schmidt, A., & Beigl, M. (2000). *Adding Some Smartness to Devices and Everyday Things*. Monterrey, USA: IEEE Workshop on Mobile Computing Systems and Applications.

Grady, P. (1999). *The age of Modularity*. Iowa: Adams and Steele Publishers.

Hodges, S., & McFarlane, D. (2005). *Radio frequency identification: technology, applications and impact*, AutoID Labs white paper.

Learmount, S., Press, M., & Cooper, R. (2000). *Design Against Crime. Report to the Design Council, Home Office and Department of Trade and Industry. ODPM (Office of the Deputy Prime Minister) (2004). Safer places. The planning system and crime prevention*. London, UK: Thomas Telford.

Robertson, D., & Ulrich, K. (1998). Planning for product platforms. *Sloan Management Review*, *39*(4), 19–31.

Simoes, R., & Sampaio, A. M. (2008). Effect of technology-driven products in the future of product design and development. In *Proceedings of RPD 2008* (paper 8048). Oliveira de Azemeis, Portugal.

Simon, A. (1969). *Science of the Artificial*. Cambridge, MA: The MIT Press.

Teixeira, F., & Piçarra, S. (2007). *Estado da arte em RFID*. Sybase. Retrieved February 20, 2008, from http://www.portalrfid.net

Ulrich, K., & Eppinger, S. (2007) *Product Design and Development*. New York: McGraw-Hill international edition.

Weiser, M. (1991). The computer for the 21st Century. *Scientific American*, *265*(3), 94–104. doi:10.1038/scientificamerican0991-94

KEY TERMS AND DEFINITIONS

Electronic Product Code (EPC): Is the basis for the information in the network. Each RFID tag has a unique EPC.

EPCglobal Network: A computer network used to share product data between trading partners.

Integral Architecture: Architecture in which the number of functions is considerably larger than the number of components, which implies that some components are involved in delivering multiple functions.

Modular Architecture: Architecture in which the number of functions is approximately similar to the number of components, where usually the interactions (as well as the interfaces) between components are relatively simple, and where each module is intended to be the main (often the only) responsible for providing a given function or sub-function of the product.

Product Architecture: The conceptualization, design, and description of a product, its components, the interfaces between components, and the relationships with internal and external entities, as they evolve over time.

Radio-Frequency Identification (RFID): Is the use of an RFID tag applied to or embedded into a product, animal, or person for the purpose of identification and tracking using radio waves.

This work was previously published in Handbook of Research on Trends in Product Design and Development: Technological and Organizational Perspectives, edited by Arlindo Silva and Ricardo Simoes, pp. 106-118, copyright 2011 by Business Science Reference (an imprint of IGI Global).

Chapter 101
Future Trends in SCM

Reza Zanjirani Farahani
Kingston University London, UK

Faraz Dadgostari
Amirkabir University of Technology, Iran

Ali Tirdad
University of British Columbia, Canada

ABSTRACT

The field of supply chain management (SCM) has experienced radical changes in its short life period. Originating from 1980 and 1990s management trends and techniques of operations management, passing process oriented and system oriented approaches in the 1990s, and now attracting the attention to behavioral approaches have all caused SCM to be largely vertiginous. So dealing with its future requires a more accurate method than common predicator fashions. Therefore, the chapter first considers SCM as a body of knowledge in which evolution is based on its theoretical foundation, and therefore, prevalent research paradigm(s), research methodological base(s) used by developers, and also real world challenges that motivate it. Consequently, the authors review current status of SCM from standpoint of the discipline's theory, its conceptualization process, and most used research methods and approaches. Then the authors will be able to use its implications to adopt an appropriate model of philosophy of knowledge for scientific change and knowledge growth of SCM. This can be used as a guide to the future of supply chain management.

AN EVOLVING CONCEPT

Supply chain Management's definition is seriously contentious. Kathawala et.al, (2003) indicates that supply chain management is defined poorly and its meaning has a high volume of variability in the minds. Also definitions provided in literature (such as New, 1997; Mentzer et al., 2001; Kauffman, 2002) shows little consensus on the definition of supply chain management. Burgess et al., 2006 have reviewed the literature of supply chain management from the standpoint of its prevalent definition status to evaluate the maturity level of the field. Their review reveals that almost 20 percent of researches use existing definitions, 15 percent develops their own definition and in half of them no definition is used. More concentration

DOI: 10.4018/978-1-4666-1945-6.ch101

shows even existing definitions which is used in some research have little consistency and there is no specifically accepted definition. Besides this considering that 10 percent of researches incrementally changes existing definition can tell us the definition of supply chain management is still open and challenging.

To realize the researchers and practitioners understanding about the concept we can investigate framing of supply chain management in their works. Its large bearing on the qualities of the way it conceptualizes will make the future trends more clear. Burgess et al. represents four categories in which Supply chain Management is conceptually framed. These four categories; “activity” category (described as a function of a process), “process” category (described as a chain of associated activities), “system” category (described as a chain of associated processes) and “other” category (described as kind of analysis which deal with, interdisciplinary, behavioral, sociological, psychological and philosophical aspects); describes four ways with which the researchers and practitioners deal supply chain management. For example assuming supply chain management an activity makes it something minor which can be treated as an operational function. On the other hand perceiving it as a process makes or prospective wider and understanding it as a system represents supply chain management as an all-embracing management framework. At last describing supply chain management considering its behavioral aspects leads to a holistic point of view. Investigating the conceptual framing which is used by researchers and organizations reveals the current construct of supply chain management and help us to understand its evolving trend. Burgess et al. show the majority of more than half of researches viewed supply chain management as some form of process. More than 20 percent framed it as a system and less than a tenth saw it as a simple activity.

Also presenting the conceptualization of supply chain management successfully depends on the construct with which we logically group supply chain management as a management concept. Constructs are about the higher order absolute immeasurable variables, but give us a clearer definition of the concepts (Nunnally, 1978, cited in Burgess et al 2006). The set of generally accepted constructs doesn’t appear to exist. In regard to the commonalities between the proposed sets of constructs for supply chain management by several researchers such as Chen et al (2004), Min et al (2001) and Tracey et al. (2004), Burgess et al conclude a set of seven constructs. This set includes “leadership” (in regard to strategic aspects of supply chain management; (see Shapiro,2004)), “inter organizational relationships”, (concentrating on the connections within the organizations), “intra organizational relationships” (concentrating on the connections between the organizations, based on joint social and economic responsibilities and interests)”logistics” (focusing on the issues which address material handling in a supply chain),”process improvement orientation” (in regard to continuous improvement of processes within a supply chain),”information system”(focusing on systems which facilitate communications within a supply chain) and “business results and outcomes”(concentrating on the economic benefits and outcomes of adopting an appropriate supply chain management aims). As other discipline within management field constructs of supply chain management can be divided to two general categories of “soft” and “hard” constructs. Soft constructs deal with behavioral and social issues of the field and hard constructs are about infrastructural and technological aspects of the field. Therefore the first three construct can be categorized as the soft ones and the others can be categorized as the hard ones.

Based on the researches done on the constructs of supply chain management and these categories by Burgess et al, even in soft construct which is expected to be concentrated on people related and behavioral issues, the research have been on technical aspects. Most of research in soft constructs is carrying on about inter-organizational relationships and almost two thirds of the research in hard construct is about process improvement orientation. Other constructs have been researched lesser but they haven't been neglected.

Every field's maturity level can be indicated by its definitions status. As described, Most of the researchers and practitioners uses newly developed definitions or develops new definitions themselves. Even existing used definitions have no convergence on a single one. A matured developed field uses few or even single definition.

Therefore and also as other researchers acknowledged (New, 1997; Cox et al., 2001; Kauffman, 2002; Quayle, 2003) supply chain management is an immature field of research and practice and should be considered as an evolving concept. Additionally the portions divided between supply chain management's constructs shows the importance of process based point of view. It's basically because of the field's Originating from 1980 and 1990's management trends and techniques of operations Management in which technical and even mathematical methods play the main role. In regards to the emphasis of many authors on the importance of system based point of view, supply chain management's social aspects, people related issues which extends cooperation and organizational trust within the chain, soft construct (concentrating on behavioral and cultural issues) can be a main pivot of maturing of the field's definitional aspects which can make it more integrated, more comprehensive and therefore more acceptable by most of researchers and practitioners. Without existence of such a definition, theoretical development of supply chain management could be deformative and even impossible and accordingly the questions such as: "how does an organization's dominant culture can form the way with which that organization uses internal rules and external standards to coordinate with the other partners within the chain.?" (Dowty et.al, 2010) Not only doesn't find any answer but even won't be on the table.

THEORETICAL DEVELOPMENT

For any field of science and practice theoretical development is a fundamental requirement for an integrated and triangular development (popper, 1961; Wacker; 1998 cited in Burgess et al). In the matter of supply chain management several disciplines claim the ownership of it and therefore they assume their theoretical basics can be used for the field's scientific and practical development. They suggest that supply chain management doesn't require an independent self motivated theory and other disciplines' theories are adoptable for theoretical development requirements. Others think an independent self motivated theory should be developed and proposed (Van Maanen; 1995). If we consider that the field's development requires a "clear line" from definition to theory and then to the research methodologies, the importance of theoretical status of the field and its trends is undeniable. Also the two way influence of development in theoretical term and development in research methodological term should be considered and therefore the research methods which are applied in the field and their paradigmatic stance should be investigated and consequently should be used aptly to scheme the future of the field.

Any discipline is based on some infrastructures which create knowledge and transfer it within a specific field. Such infrastructures should contain defined theories, applicable research methods and practicable tools. While the dominant discipline which is leading supply chain management's research and practice is Operations Management, a multitude of the other disciplines are emerging to contribute in the future of the field (Sachan et al). It means that these disciplines' researchers tend to use theoretical basics and research methods of their disciplines to develop the field.

Burgess et al recognize nine disciplines relevant to supply chain management including, "Marketing", "Purchasing", "Strategy", "Finance/Economics", "Sociology/Psychology", "Information", "Logistics", "Operations Management" and "Others". Based on their findings, operations management, strategy and purchasing are the disciplines which are accomplishing most of the research activities (and therefore have the biggest portion in the developing of supply chain management, based on the theoretical basis of their discipline) in the field. Also in this research have been shown that multidisciplinary researches have small role in the developing the field's body of knowledge. Also as Voss et al (2002) mention, the supply chain management's body of knowledge is developed leading by practitioners and is following by the researchers. Accordingly low presence of researches focusing on the soft type, psychological and sociological aspects of the field and the importance of them in practitioners' success in the business shows future potentials of these disciplines' contributions in the supply chain management growth. But to understand the reasons causing the incoherence of researches has done till now we should back to the theoretical concerns and research methodological stands of the field.

As the same as the research methods which have applied in the field, practitioners and researchers have used the same point of view about assumed theoretical basics in the research and practice. Most used theories respectively are strategic management related theories (about 40 percent), transaction cost theory of economics (about 30 percent), and psychological/sociological related theories (about 10 percent). But 25 percent of the researches don't use discernable theory. Also there is no supply chain management specific theory, used in any published research. Furthermore no multi-theoretic based research has accomplished yet (Burgess et al).

Despite the multidisciplinary basics of supply chain management and usage of different type theories by researchers and practitioners it seems in future there should be trends to fill these gaps. Thus the researchers and practitioners need to develop and use supply chain management specific theories, multi-theoretic basics and also multidisciplinary approaches. Otherwise these gaps, current incoherency can inhibit the field's regular development and can't logically be persuaded. Also theoretical contributions are more descriptive than other types. This indicates that supply chain management is an immature field from standpoint of the theoretical basics. Additionally while theoretical contributions aren't developing the conceptual aspects of supply chain management by theoretical validation, theoretical development and theoretical refinement, supply chain management can't extend its theoretical basics and therefore can't face emerging challenges. Accordingly the field's conceptualization should use more sophisticated and higher-level theoretical contributions to encounter the face the challenges.

Burgess et al shows that supply chain management needs a Meta-Theory to overcome the theoretical difficulties. Meta-theory is a "theory of theories" which firstly developed by Tsoukas 1993 for Management. It creates an appropriate theoretical foundation for the fields which have developed based on different theoretical basics, within various disciplines and belongs to distinct traditions. It is suited for the areas which epistemological and ontological diversity is high and single theory can't serve the fields development.

Considering the above and based on Tsoukas, an appropriately developed Meta-Theory can merge the ontological and epistemological principles, which can explain the foundations of the field and therefore the existing and possible body of Knowledge within it.

PROSPECTIVE BODY OF KNOWLEDGE

What mentioned before shows supply chain management has conceptualized as a "science" and positivist paradigm is dominant in its theoretical and research methodological aspects. Relative lack of non-positivist approaches and absence of Mixed-Multi-paradigmatic research methods can prevent triangulation to Body of Knowledge

Development in the field. These issues can be caused by practical difficulties in extending these type methods and also given the wide opposing assumptions that are made (Burrell et.al, 1979; cited in Burgess et.al).

Considering the above, development of the Body of Knowledge of supply chain management should be fill the mentioned gaps and overcome the preventing difficulties. From standpoint of the Philosophy of Knowledge the appropriate way to describe a body of knowledge is considering a fundamental theory is phenomenology. Phenomenology in Husserl's perspective basically is related to the systematic reflection on and analysis of the structures of the consciousness and how a phenomenon appears in acts of consciousness. It seeks through systematic reflection to determine the fundamental attributes and structures of conscious experience and also consciousness itself (Husserl, 1973). He believed that phenomenology provides a foundation for scientific knowledge, and ground philosophy as a "rigorous science" of measurable perception. In this manner Lakatos proposed that there is something named "research program" with which researchers study a phenomenon based on different theories and these theories operates within it. In the Lakatosian research program a "hard core" that contains some theories which are informed positively provide some foretelling quality about a phenomenon.

On the other side in this research program there is a "protective belt" which tends to coalesce with hard core and progressively helps the field's growth by the discovery of astonishing novel facts, development of new experimental Methods, more accurate predictions and also by introducing new and applicable aspects of definitive and theoretic point of views to the field. Also the researchers who are engaged in the program try to prevent the hard core from falsification endeavors behind the hypotheses of the protective belt. This provides us a Model to understand how the knowledge of a field of science and practice grows.

Based on finding of Burgess et al hard core in the supply chain management's body of knowledge is based on "operations management, manufacturing, analytical school, process and positivist paradigm". Also the protective belt contains several other activities which aren't defined clearly and need to more attention and can be the basics of future trends. Behavioral approaches concentrated on Psycho-social dimensions of the field and also multi-disciplinary and Multi-theoretical aspects should be considered to describe the most realistic identity of protective belt related activities. Based on it and applying Lakatosian approach we can aptly describe how knowledge growth appears to be conducted in the field facing the emerging challenges. It provides a logically framed analytical way to perceive the way which the SCM's body of knowledge is evolving rather than simply using common predicator fashions.

Accordingly supply chain management is a young immature field of research and practice which its evolving path should make a serious conceptual growth based on a defined meta-theory which can overcome theoretical, research methodological and other knowledge empowering related basics difficulties and also can fill gaps caused by incoherencies and lack of integrated clear path of insight in the research methodological stand point.

This exponentially growing field should develop and apply research methods, theoretical basics and methodological stand points of the protective belt to help to reduce dissociations in the field and interacting with novel contents and methodologies, along with staying connected with the theoretical core of supply chain management in contents and methodologies.

FUTURE OF THE REAL-WORLD APPLICATIONS OF SCM IN THE INDUSTRY

We are going to talk about different research directions for SCM which have been interested for industry and business and still there are lots of works to do in these areas. We will go through the following topics: New Public-Private Partnership Trend, supply chain design and redesign, supply chain coordination, supply chain risk management, and at the end we are going to provide some case studies from real world problems.

New Public-Private Partnership Trend

In general we tend to optimize the supply chain. In a stable set of supply chain network it is possible to model and optimize the network and we can have a supply chain coordinator. But the current environment of dynamic demand-driven supply networks, most of time leads to emerge of the temporary supply chains. The lifespan of these supply chains directly depends on the market opportunity. In such dynamic networks the speed of execution is usually more important than the finding the optimum of the network and the speed of execution is directly related to the coordination of partner. The excellent examples of the extreme dynamic environment are war and disaster relief (Kumar, 2001). For a special case we can mention the weak performance of the governmental agencies in Hurricane Katrina in States, in delivery of supplies and equipment. Wal-Mart was the fastest organization which handled delivery of the goods like food, water and etc. to the victims (Dimitruk, 2005).

So the old-school model of the government as the "Lord of the Supply Chain Network" is not working anymore in this environment. That's why the tendency observed in the developed countries to reconsider the pubic-private partnerships.

The other concern about reconsidering the public-private relationship emerged after September 11th, 2001. The SC managers challenged to adjust their relationship with others in the network. The security and being prepare for dealing with problems after the terrorist attacks are the main challenges that effect on the public-private relationships (Sheffi, 2001).

Designing and Redesigning the Supply Chain Networks

Supply chain configuration for mass customization and supply chain quality modeling are two major concepts for designing a supply chain networks. We will go briefly to the main works which talk about these two topics.

Salvador et.al, (2004) discuss about the problem of supply-chain configuration for mass customization and suggest an answer. From their work, other questions arose for the further research, for example "do modular products need a modular supply chain network or non-modular ones work for them either?"

Batson and Mc Gough (2007) talk about how the network model can enhance the quality in manufacturing supply chains. They outlined these future research topics in supply chain quality modeling:

- **Multi-company adaptations of Experimental:** For reaching to the better quality (closer to the real optimum of the network) the planning, coordination, execution, and data analysis should perform across multiple levels in the supply chain network. This concept arise new research topic to combine the design of experiment (DOE) and supply chain management.
- **Performance analysis:** Supply chain performance analysis (e.g. multi-level tolerance analysis) should perform before the production. Cost and quality are regularly achieved worse than the optimum values in multi-level environments. How can these be avoided? That's another topic for further work.
- **Statistical process control:** Woodall and Montgomery (1997), they discuss that the quality improvement (QI) is a challenging work in the multi-level environments like supply chain networks.

By redesigning the supply chain network, we mean the investigation of process efficiency and process effectiveness. Redesign can result in increasing the operational efficiency or effectiveness of the supply chain network (Hewitt, 1994). Moreover, rapidly changing business situations make it necessary to redesign the supply chain management time to time. The example of the rapidly changing supply chain networks are large networks; because small change in one of the suppliers can results in a huge changes in the whole network (bullwhip effect) (Van der Vorst,2002). Van der Vorst et.al (2002) claim that the uncertainty in the supply chain network is the one of critical criteria in redesigning. So based on this claim, they provide a research approach to supply chain redesign and declare that this approach is applicable for large chain network either.

Service Supply Chain

The service operations should be considered differently from manufacturing products; because they have different characteristics which you cannot find in manufacturing materials. The service operations are intangible, labor intensive, heterogeneous, we are not able to store them or transport them and the quality of a service is hard to determine in comparison to the manufacturing products (Slack, 2001).

So based on the different, the service supply chains inhere the different from manufacturing supply chains. In designing service supply chains effectively, we can still use the experience and guidance from manufacturing supply chain. In fact, manufacturing supply changes are more understandable, because they are tangible and there is a rich literature on it over the past 20 years.

Akkermans et.al, (2003) show that part of the available knowledge from manufacturing supply chains can be use for service supply chains as well, with some remarkable exceptions and some additions. These exceptions come from this fact that services cannot be stock but can be ordered. So first obvious result is no inventory level but they have workload levels. The other result is the known measures of inventory do not apply for services.

Supply Chain Management and Facility Location

Context of supply chain management contains problems which can be discussed with the facility location problems structure. Using facility location concepts and models to address supply chain management problems should have characteristics which meet needs of supply chain management problems. Considering this assumption facility location can extend its future in this context in different directions. Developing models which can consider uncertainties of supply chain management, which can address multi-layer networks and stochastic constraints and Modeling parameters such as researches of Miranda et.al. (2008) will extend applicability of the facility location problems in this context.

Concentrating on vehicle routing problems along with the transportation systems that are used trough the chain such as works of lee et.al, (2008), aksen et.al, (2008) and many other relevant tactical/operational systematic decisions such as procurement, reverse activities, postponement decisions and backorders are getting more critical. Also considerable simplifications have been done on the models to make them efficiently solvable. Developments in computational infrastructures lead to more realistic and more complicated modeling contributions and more accurate solution methods. The coming capabilities of the optimization applications and expected computation power of coming hardware makes the future models more complex and comprehensive and will enhance the possibility of solving real-world problems of supply chain management. Modeling procedures relevant to facility location tend to use multi-objective type models (such as Farahani et.al., 2007; Du et.al, 2008 and Mansouri et.al, 2010) along with new type of objective functions such as maximizing the potential return of the investments (see sheu, 2003; ReVelle et.al., 2008; Mirchandani et.al., 1990) or profit oriented objective functions to add the idea of revenue management to supply chain management context. Also lack of attention to planning for distribution in supply chain management can cause serious risks. Accordingly, it is essential to consider disruptions during the modeling of supply chains so that they perform well even after a disruption. Resulting models made substantially more reliable with optimum additional investments (Snyder et. al., 2005).

The integration of strategic and tactical aspects of supply chain management is growing and therefore the application and role of facility location related concepts of features are getting more important. Considering various aspects of supply chain management simultaneously, modeling real world problems makes the body of Knowledge of the facility location more applicable in the future of strategic supply chain management planning.

Reverse Supply Chain Management

Recent trends shows the newly emerging practical issues which are getting more serious and the researchers and practitioners should consider them to develop appropriate managing approach in the supply chains. Reverse logistics is one of those and its applications are getting broader and more unavoidable. One of the most effective reasons to this is the environmental law constraints which are being imposed in the word with fast growing trend and is being considered by practitioners and researchers seriously (see Jalali Naini et.al, 2010). Accordingly it leads to the need for safe return of products from the end of the chain or even from the middle of it. Therefore supply chain management should now consider both forward and return flows of products, parts, subassemblies, scrap and containers. These completely new spectrum of goods should be managed at where was considered the end of the chain. There may be occasions for most of the chains when the goods they have produced and supplied to the market should be returned through the reverse chain function. These may include the products which have failed or become absolute but aren't useless yet, also the unsold or recalled products should be brought back to the chain and at last the containers which should be brought back to be used again. These examples shows the importance of the reverse supply chain management and the efficiency which can be lost by ignoring it. (Helms et.al, 2006; Du et.al, 2008)

Broader economic perspective illustrates the beneficial of reverse logistic systems such as recycling systems and infrastructure for recyclable materials to communities. Of the many practical factors technical issues pertaining to reverse supply chain management such as facility locations, distribution points, scheduling issues and operational inventory levels, have been investigated widely and should be investigated in the future further. In fact these are important issues in the design of any reverse logistics network. Organizations are feeling the pressures from the regulatory or from the social opinions. Also competiveness is another reason which stimulates the organizations to assign optimum recourses to manage reverse logistics. Therefore this field is becoming prominent in the research and practice of supply chain management. Considering the robust nature of the researches which have been conducted yet, the researchers have opportunity to develop the theoretical foundations and also the technical aspects of the area. They can focus on the final stages of organizational management. Also theoretical and methodological standards should be developed to describe the phenomena, advance the practice of reverse logistics and as a result broaden the methodologies to improve incorporation within the supply chain. (Meade et.al, 2007)

Reducing Bullwhip Effect in Supply Chains

Small fluctuations in consumer demand extremely get larger when they go up through the supply chain from retailers to main suppliers. In general the variance of consumer demands sometimes can get larger than the demands; this is called “Bullwhip Effect” (Lee et.al, 1997). There are four main reasons that the Bullwhip Effect happens (Lee et.al, 1997):

- Consumers’ interpreting demands,
- order in batches,
- promotions which changes the artificial demand (not the real demands) and
- Shortage in supply which results in emergence of the demand artificially either.

There are several papers who worked on reducing the Bullwhip Effect in supply chains. Warburton et.al, (2004) provide ordering policies to reduce the Bullwhip Effect and improve inventory management. Lee et.al, (1997) suggest the followings for reducing the Bullwhip Effect: Choosing the combination of sell according to data, sharing the inventory status information between the different levels in supply chain, ordering with coordination in advance and finally using simplified pricing rules. A number of methods of mitigating Bullwhip Effect have been reviewed by Towill (1997). These methods work in the real world problems properly. They showed that the Bullwhip Effect can be reduced via Business Process Reengineering (BPR) up to 50%.

Vendor Management Inventory (VMI) and Reengineering in Supply Chains

Vendor Management Inventory (VMI) is a strategy which allows the retailer to manage the customer’s stock on his own (the supplier does not get involved with this responsibility). VMI has become more popular as retailers get more powerful recently. The great example is Wal-Mart which in the grocery sector in the last 15 years has VMI strategy and has a successful performance (Disney et.al, 2003). Moreover, we can say somehow recently the necessary ICT has become available in acceptable cost to enable this strategy. However it has been shown that it can be performed via fax or emails and spreadsheets (Holmstrom, 1998). VMI have been showed up differently in the literature. Quick response (QR), synchronized consumer response (SCR) and centralized inventory management (CIM) are the applications of VIM strategy (Disney et.al, 2003).

Supply change reengineering usually means improvement in processes of the supply chain in terms of the cost and quality of service in general. This need of improvement can arise from the competition, increasing in customer expectation or etc (Swaminathan et.al, 1998). One of the main drivers for reengineering the supply chain is to reduce product life-span. Since the early 80’s supply chain reengineering strategies has been performed by companies. Before that, the idea of reengineering had existed in the supply chain but not in a proper way. Just in time (JIT) strategy was the one of the closest idea to the reengineering in the early 1980s (Naylor et.al, 1999). Nowadays, by extremely competitive environment, tendency to implement the reengineering more frequently for increasing the quality of service and lowering the costs has a upward direction.

Multi-Channel Cooperation in Supply Chains

By improving the supply chain concept every day, there is the need for companies to look outside of box and look for opportunities to collaborate and coordinate with partners. This helps them to have a more efficient supply chain for demands with high variances. However, these opportunities lead to new challenges and concerns; like huge supply chain scale and possibility of conflicting between the sides of the cooperation (Balakrishnan et.al, 2004).

Providing products or services to consumers via different channels of suppliers or even retailers is called multi-channel supply chain. It leads to increase of levels of quality of service, or on the other hand it results in challenging issues like complexity and more responsibility for supply chain managers. In other words, the more channel options we have, the more complex supply chain will emerge. Synchronizing the channel of supply chain in the way that the consumers' need will meet independent of which channel is chosen to serve the consumer, needs the latest ICT technology and management experts. Radio frequency identification (RFID) is one of the technologies that can help multi-channel supply chain management a lot (Rosenbloom, 2006). Journal of "Industrial Marketing Management" Volume 36, issue 1 (2007), have a good collection of papers about the multi-channel coordination supply chains, the problem to handle them, some approaches to obtain the optimal channel and etc. Moreover, Stadtler (2009), provides a review on collaborative planning in supply chain networks.

Supply Chains and Risk Management

The business environment is expanding every day and as a result, rougher and more frequent natural and human-made disasters make supply chains more vulnerable (Wagner et.al,2010). Today supply chains are getting more complex and therefore to have an efficient supply chain and also to keep the market consumers satisfy, we need to apply risk management concepts and Models. In other words, it is necessary for the supply chain managers to identify, evaluate, sort, and manage any kind of risk which threat the network.

The dynamicity of a supply chains means that the decision model for risk management should reflect the interaction and relationship between managerial elements and factors of the risk (Xia et.al, 2011). Risk consists of uncertain outcomes with known possibilities. The sources of risks can be from inside or outside of supply chain. The role of risk management is to recognize the risks who endanger the network and find the way to minimize the effect of the on the outcomes of supply chain (Faisal et.al, 2006). Mainly there are two types of risks for the supply chain networks; disruptions and delays (Craighead et.al, 2007).

In the today's business environment, survival is not an issue for companies against each other; it is supply chains' issue. So in such environment, the importance of risk management can be felt increasingly (Craighead et.al, 2007).

Real-Time Supply Chain Management

Collaboration within the supply chain in the current market of competitiveness and dynamicity is critical. Relationship among supply chain competence and supply chain agility in the current and emerging global supply chains are undeniable (Ngai et.al, 2010). New technologies such as radio frequency identification (RFID), electronic sensors which can be used to control material flows, GPS receivers and the Navigation systems are being used to process the information within the chain in the real-Time manner. These advances in IT and ICT make us capable to address and solve new decision making problems which can be cross-enterprise or even cross-functional and improve the firms performance by improving the supply chains agility (Ngai et.al, 2010). From standpoint of mathematical modeling of supply chain management problems, Real-Time Modeling can be used instead of dynamic or Stochastic Modeling. In fact real-time management aims to empower operations research capability to Model uncertainty and to solve dynamic decision making problems. Therefore real-time flow of information through the supply chain makes us capable of modeling new type of problems in which incomplete or unknown data are be known concurrently with the solution process and as the result we can make more accurate decisions on-line. This simply can improve efficiency by enhancement of traceability and visibility of products and processes, enhancements in the efficiency of the processes, improvements in the accuracy of information flow, reductions in the delays, Reductions in the losses of the inventory and also Facilitation of management through real-time information (sarac et.al, 2010).

Newly developed technologies and also the companies' necessity to manage their operations in the real-time manner have made the research and practice of the real-time supply chain management flourishing. Dynamic problems as Psaraftis (1995) indicates differ from the static problems at least in 12 aspects and therefore, considering these aspects not only affect the strategic policies but even can improve simple decision making procedures.

Recent works on dynamic routing problems (Larsen et.al.,2007), dynamic pricing in the real time supply chain management (Burke et.al.,2008), modeling dynamic demand in supply chains (Sodhi et.al,2009; Toktas-Palut et.al,2011) and even on integrated framework to develop RFID technology in the logistics and supply chain management (Lin,2009) show the vast trend of researchers and practitioners to develop the theoretical foundations and also the applications of the concepts, the tools and the technologies which can be used in the real-time supply chain management.

Supply Chain Management and Scale Economies

Considering economies of scale in designing a supply chain is critical. If someone does not consider scale economies in supply chain, it can lead to significantly higher costs rather than the optimal design. One of the practical examples of importance of economies of scales is provided by Baumgartner et.al, (2009). Supply chain management's complexity (Bozarth et.al, 2009) complicates finding suitable measures for supply chain. In the other words supply chain performance evaluating is complex, because this is a crossing process containing various factors cooperating to achieve the logistical and strategic objectives (Estampe et.al, 2010). Cohen et.al, (1990) present a review of the performance measures and provides a framework for the selection of measures for manufacturing supply chains.

In practice the performance of the supply chains is measured by the customers' level of satisfaction and the expenditures. These evaluations can be especially important in the case, where supply networks are being considered a critical factor of the organizations' success (Estampe et.al, 2010).

PROBLEM MODELING

Most of the time, supply chain models are really huge; because they should consider the large number of materials, suppliers, retailers and etc. and they require a huge data collection. If we want to categorize the supply chain designs by modeling approaches, it is possible to categorize them in four general classes: (1) deterministic models (2) stochastic models (3) economic models, and (4) simulation models (Beamon, 1998). We are going to review the literature briefly for each of these classes.

Deterministic Supply Chain Models

In the deterministic models, all the parameters of the problem are known and there is no uncertainty in their value. Like most of the other problems, the very first approach for this type of problems is heuristic algorithms. Williams (1981) provides seven heuristics with a minimum cost of production and distribution as the objective function. The only constraint considered in this model was satisfaction of the consumers' demands.

Dynamic programming was the next approach for solving deterministic models (Williams, 1983). In this model additional to the above objective function, holding cost is considered either at each node of supply chain network.

A mixed integer, non-linear mathematical model, is provided by Cohen and Lee (1989). The objective function was maximization the total profit for all the participants in the supply chain network. They consider two types of constraints: boundary conditions which talk about the maximum of the available resources and limitation of demands which determine that each demand cannot be more and less than certain amount. Cohen and Moon (1990) extend the previous work and survey the effect of different types of factors on the total cost of supply chain network. In this extended model, multiple products are acceptable and transportation cost is considered either. The objective function was the somehow same as before, but they consider more constraints including requirement of materials at each node, capacity and assignment limitations. (see Chan et.al, 2011)

A Two-phase approach is represented by Newhart, et. al. (1993). This approach first minimizes the number of product classes in inventory using a hybrid method of mathematical and heuristic model and then it provides the alternatives with the minimum value for required safety stock in a feasible way. Finally we should choose the option with a minimum cost, obviously.

Generally, it seems that using mathematical programming have been become more popular after this point. Arntzen et. al. (1995) design a mixed integer model considering multiplicity in most of the features of the network, e.g. products, transportation options, levels of the network and etc. Voudouris (1996) builds a mathematical model with considering efficiency and flexibility as part of objective function to maximize. Chan and Chung (2004) provide multi-criterion genetic optimization technique using analytical hierarchy process for a large supply chain network.

Stochastic Supply Chain Models

Stochastic supply chains models include models with at least one unknown parameter. Usually in these cases it is considered that the unknown parameter has a specific probability distribution for its possible values. First stochastic parameter which comes to the mind in supply chain problems is demand. In Beamon (1998), lots of papers which discuss about stochastic demands are provided in literature review. For example, Lee, et. al. (1997) design stochastic mathematical models which can reveal the effect of different factors of the Bullwhip effect.

We should not forget that the critical parameters of most of supply chains, like demands, resources and prices have uncertainty, somehow. But the problem is that most of the works in the literature consider these parameters as a deterministic parameter and some stochastic mathematical models are not applicable in the realistic scale of real world problems (Santoso et. al., 2005). So it is time to consider this aspect in the future researches. Santoso et. al. develop a stochastic mathematical model for supply chain networks and provide a solution for that by Bender's decomposition method.

Goh et. al. (2007) provides a two-stage stochastic model for the multi-stage supply chain network problems with objective of maximizing the profits and minimizing the risks. They mention the large scale supply chain network of this type for the future research. Also Defaults of companies in the multi-stage supply chain networks have been modeled in Mizgier et.al, (2010) in which the stochastic elements of supply chains are introduced and how local events can affect the global performance of the chain has been shown.

Also uncertainty can be considered and modeled in other ways instead of stochastic optimization paradigm. Especially strategic level supply chain management deals with uncertainty with modeling the relationships of strategic level decisions and selected product, operational and demand variables. (see Wanke et. al.2003)

Economic Supply Chain Models

By economic models we address to the models which use mostly game theory concepts for modeling the supply chains. The two sides of the games are usually buyer and supplier (Beamon, 1998). The evolution of the supply chains in the recent decade reveal important challenge to supply chain managers; every business process contains several decentralized organizations which final decisions of these different firms impact profit of other firms. It means that the entities in the supply chain network are like the players in a game. So for being able to model and analyze strategies in this environment, game theory is a logical choice (Nagarajan and Sosic, 2006). Cachon and Netessine (2006) is a valuable reference for the application of game theory in supply chain analysis.

Simulation Supply Chain Models

Simulation plays a significant role in either modeling the supply chain network or in the validation of the mathematical models for the supply chains. Terzi and Cavalieri (2004) provide a great survey on the application of the simulation in the supply chain network over the 80 articles. It can be a suitable reference for anyone who interested.

CONCLUSION

It looks there are various ways for future of supply chain management. First of all, the models are going to get more complicated for lager supply chains. Because the aim of SCM is going to hit the larger targets and analyze more inclusive networks. Therefore they are more difficult to solve. Hence the need of faster and more efficient solution methods is tangible. Another topic is the modeling of supply chain networks considering reassigning demands or re-routing flow of materials after a disruption and their costs. Moreover, risk mitigation is important concern in the nowadays competitive business environment. In the competitive environment, most of time it is enough to make sure that the objectives of our competitors are not better than us. This idea rings the bell of using game theoretic models to approach supply chain networks. Speaking of the objective of supply chain networks is mostly about the cost, But there are especial cases in the public services such as emergency post-disaster supplies, which their reliability have a significant role. This can bring a challenge to modeling and applying this kind of problems (Snyder, et. al, 2005). The very last case is one of the applications of the mixture of facility location and supply chain management. Melo et.al, (2008) provide a good review about this topic. They mentions that stochastic SCM is one of the areas in SCM which needs an intensive research work on it. The other hot topics in SCM mixed with facility location problems are the cases with procurement, routing and the choice of transportation modes.

Another concern which is not considered properly in the supply chain management researches is formulation of regulation to societal and ecological concerns. These frameworks are called environmental or green supply chain management. Srivastava (2007) makes a review on this topic and provides some directions for future research in this area. Shen (2007) discusses about the other interesting topics of the future researches. Considering limitation for the capacity of demand, developing new algorithms to efficiently solve the large-scale networks, game theoretic approach to competitive supply chains and including robustness, reliability and risk management in designing supply chain networks can be the ways to the future works. Aside from these, definitive, theoretic and methodological difficulties of body of knowledge of supply chain management are getting more challenging and therefore future researchers and practitioners are to fund new basics to overcome the theoretical shortages and therefore to develop supply chain management as it becomes able to face the time concerns and challenges.

Extensive challenges are emerging as it mentioned before. Different applications of supply chain management are getting more serious roles in the future of the industries or even in the future of the world. Economic challenges as environmental threats such as global warming are making supply chain management a leading state of the art in the humans' hands which can help to build a better life for us and also the coming generations.

REFERENCES

Akkermans, H., & Vos, B. (2003). Amplification in service supply chains: An exploratory case study from the telecom industry. *Production and Operations Management*, *12*(2), 204–223. doi:10.1111/j.1937-5956.2003.tb00501.x

Aksen, D., & Altinkemer, K. (2008). A location-routing problem for the conversion to the ''click-and mortar'' retailing: The static case. *European Journal of Operational Research*, *186*, 554–575. doi:10.1016/j.ejor.2007.01.048

Arntzen, B. C., Brown, G. G., Harrison, T. P., & Trafton, L. L. (1995). Global supply chain management at digital equipment corporation. *Interfaces*, *25*, 69–93. doi:10.1287/inte.25.1.69

Balakrishnan, A., & Geunes, J. (2004). Collaboration and coordination in supply chain management and e-commerce. *Production and Operation Management Society*, *13*(1), 1–2. doi:10.1111/j.1937-5956.2004.tb00140.x

Batson, R. G., & Gough, K. D. (2007). A new direction in quality engineering: supply chain quality modeling. *International Journal of Production Research*, *45*(23), 5455–5464. doi:10.1080/00207540701325140

Baumgartner, K., & Thonemann, U. W. (2009). *Supply chain design considering economies of scale and transport frequencies*. Technical report.

Beamon, B. M. (1998). Supply chain design and analysis: Models and methods. *International Journal of Production Economics*, *55*(3), 281–294. doi:10.1016/S0925-5273(98)00079-6

Bozarth, C. C., Warsing, D. P., Flynn, B. B., & Flynn, E. J. (2009). The impact of supply chain complexity on manufacturing plant performance. *Journal of Operations Management*, *27*, 78–93. doi:10.1016/j.jom.2008.07.003

Burgess, K., Singh, P. J., & Koroglu, R. (2006). Supply chain management: A structured literature review and implications for future research. *International Journal of Operations & Production Management*, *26*(7), 703–729. doi:10.1108/01443570610672202

Burke, D. A., Brown, K. N., Tarim, S. A., & Hnich, B. (2008). Learning market prices in real-time supply chain management. *Computers & Operations Research*, *35*(11). doi:10.1016/j.cor.2007.01.021

Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organizational analysis*. London, UK: Heinemann Books.

Cachon, G., & Netessine, S. (2004). Game theory in supply chain analysis. In Simchi-Levi, D., Wu, S. D., & Shen, M. (Eds.), *Supply chain analysis in the e-business era*. Kluwer.

Chan, F. T. S., & Chung, S. H. (2004). A multi-criterion genetic algorithm for order distribution in a demand driven supply chain. *International Journal of Computer Integrated Manufacturing*, *17*(4), 339–351. doi:10.1080/09511920310001617022

Chan, F. T. S., & Zhang, T. (2011). The impact of collaborative transportation management on supply chain performance: A simulation approach. *Expert Systems with Applications*, *38*, 2319–2329. doi:10.1016/j.eswa.2010.08.020

Cohen, M. A., & Lee, H. L. (1989). Resource deployment analysis of global manufacturing and distribution networks. *Journal of Manufacturing and Operations Management*, *2*, 81–104.

Cohen, M. A., & Moon, S. (1990). Impact of production scale economies, manufacturing complexity, and transportation costs on supply chain facility networks. *Journal of Manufacturing and Operations Management*, *3*, 269–292.

Cox, A., Sanderson, J., & Watson, G. (2001). Supply chains and power regimes: Toward an analytic framework for managing extended networks of buyer and supplier relationships. *Journal of Supply Chain Management*, *37*(2), 28–35. doi:10.1111/j.1745-493X.2001.tb00097.x

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., & Handfield, R. B. (2007). Conceptual note - The severity of supply chain disruptions: Design characteristics and mitigation capabilities. *Decision Sciences*, *38*(1), 131–156. doi:10.1111/j.1540-5915.2007.00151.x

Dimitruk, P. A. (2005). *Three keys to supply chain management in times of disaster*. Heathcare Purchasing News.

Disney, S. M., & Towill, D. R. (2003). The effect of vendor managed inventory (VMI) dynamics on the bullwhip effect in supply chains. *International Journal of Production Economics*, *85*, 199–215. doi:10.1016/S0925-5273(03)00110-5

Dowty, R. A., & Wallace, W. A. (2009). Implications of organizational culture for supply chain disruption and restoration. *International Journal of Production Economics*, *126*, 57–65. doi:10.1016/j.ijpe.2009.10.024

Du, F., & Evans, G. W. (2008). A bi-objective reverse logistics network analysis for post-sale service. *Computers & Operations Research*, *35*, 2617–2634. doi:10.1016/j.cor.2006.12.020

Estampe, D., Lamouri, S., Paris, J. L., & Brahim-Djelloul, S. (2010). (in press). A framework for analyzing supply chain performance evaluation models. *International Journal of Production Economics*..doi:10.1016/j.ijpe.2010.11.024

Faisal, M., Banwet, D., & Shankar, R. (2006). Supply chain risk mitigation: Modeling the enablers. *Business Process Management Journal*, *12*(4), 535–552. doi:10.1108/14637150610678113

Farahani, R. Z., & Asgari, N. (2007). Combination of MCDM and covering techniques in a hierarchical model for facility location: A case study. *European Journal of Operational Research*, *176*, 1839–1858. doi:10.1016/j.ejor.2005.10.039

Goh, M., Lim, J. Y. S., & Meng, F. (2007). A stochastic model for risk management in global supply chain network. *European Journal of Operational Research*, *182*, 164–173. doi:10.1016/j.ejor.2006.08.028

Helms, M. M., & Cengage, G. (2006). Reverse supply chain logistics. *Enotes Encyclopedia of Management.* Retrieved from http://www.enotes.com/management-encyclopedia/reverse-supply-chain-logistics

Hewitt, F. (1994). Supply chain redesign. *International Journal of Logistics Management*, *5*(2), 1–9. doi:10.1108/09574099410805162

Holmstrom, J. (1998). Business process innovation in the supply chain: A case study of implementing vendor managed inventory. *European Journal of Purchasing and Supply Management*, *4*, 127–131. doi:10.1016/S0969-7012(97)00028-2

Jalali Naini, S. G., Aliahmadi, A. R., & Jafari-Eskandari, M. (2010). Designing a mixed performance measurement system for environmental supply chain management using evolutionary game theory and balanced scorecard: A case study of an auto industry supply chain. *Resources, Conservation and Recycling*, *55*(6). doi:.doi:10.1016/j.resconrec.2010.10.008

Kathawala, Y. K., & Abdou, K. (2003). Supply chain evaluation in the service industry: A framework development compared to manufacturing. *Managerial Auditing Journal*, *18*(2), 140–149. doi:10.1108/02686900310455137

Kauffman, R. G. (2002). Supply management: What's in a name? Or, do we know who we are? *The Journal of Supply Chain Management, Fall,* 46-50.

Kumar, K. (2001). Technology for supporting supply chain management. *Communications of the ACM*, *44*(6), 58–61. doi:10.1145/376134.376165

Lee, D. H., & Dong, M. (2008). A heuristic approach to logistics network design for end-of-lease computer products recovery. *Transportation Research Part E, Logistics and Transportation Review*, *44*, 455–474. doi:10.1016/j.tre.2006.11.003

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distribution in a supply chain: The bullwhip effect. *Management Science*, *43*(4), 546–558. doi:10.1287/mnsc.43.4.546

Lin, L. C. (2009). An integrated framework for the development of radio frequency identification technology in the logistics and supply chain management. *Computers & Industrial Engineering*, *57*, 832–842. doi:10.1016/j.cie.2009.02.010

Mansouri, S. A., Gallear, D., & Askariazad, M. H. (2010). Decision support for build-to-order supply chain management through multi-objective optimization. *International Journal of Production Economics*, *135*(1). doi:.doi:10.1016/j.ijpe.2010.11.016

Meade, L., Sarkis, J., & Presley, A. (2007). The theory and practice of reverse logistics. *International Journal of Logistics Systems and Management*, *3*, 56–84. doi:10.1504/IJLSM.2007.012070

Melo, M. T., Nickel, S., & Saldanha-da-Gama, F. (2009). Facility location and supply chain management-A review. *European Journal of Operational Research*, *196*(2), 401–412. doi:10.1016/j.ejor.2008.05.007

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, *22*(2), 1–25. doi:10.1002/j.2158-1592.2001.tb00001.x

Miranda, P. A., & Garrido, R. A. (2008). Valid inequalities for Lagrangian relaxation in an inventory location problem with stochastic capacity. *Transportation Research Part E, Logistics and Transportation Review*, *44*, 47–65. doi:10.1016/j.tre.2006.04.002

Mirchandani, P. B., & Francis, R. L. (Eds.). (1990). *Discrete location theory*. New York, NY: Wiley.

Mizgier, K. J., Wagner, S. M., & Holyst, J. A. (2010). Modeling defaults of companies in multi-stage supply chain networks. *International Journal of Production Economics*, *135*(1). doi:.doi:10.1016/j.ijpe.2010.09.022

Nagarajan, M., & Sosic, G. (2006). Game-theoretic analysis of cooperation among supply chain agents: Review and extensions. *European Journal of Operational Research*, *187*, 719–745. doi:10.1016/j.ejor.2006.05.045

Naylor, J. B., Naim, M. N., & Berry, D. (1999). Leagility: Integrating the lean and agile manufacturing paradigms in the total supply chain. *International Journal of Production Economics*, *62*, 107–118. doi:10.1016/S0925-5273(98)00223-0

New, S. J. (1997). The scope of supply chain management research. *Supply Chain Management*, *2*(1), 15–22. doi:10.1108/13598549710156321

Newhart, D. D., Stott, K. L., & Vasko, F. J. (1993). Consolidating product sizes to minimize inventory levels for a multi-stage production and distribution systems. *The Journal of the Operational Research Society*, *44*(7), 637–644.

Ngai, E. W. T., Chau, D. C. K., & Chan, T. L. A. (2010). Information technology, operational, and management competencies for supply chain agility: Findings from case studies. *The Journal of Strategic Information Systems*, *20*(3). doi:.doi:10.1016/j.jsis.2010.11.002

Popper, K. R. (1961). *The logic of scientific discovery*. New York, NY: Science Editions.

Quayle, M. (2003). A study of supply chain management practices in UK industrial SMEs. *Supply Chain Management: An International Journal*, *8*(1), 79–86. doi:10.1108/13598540310463387

ReVelle, C. S., Eiselt, H. A., & Daskin, M. S. (2008). A bibliography for some fundamental problem categories in discrete location science. *European Journal of Operational Research*, *184*, 817–848. doi:10.1016/j.ejor.2006.12.044

Rosenbloom, B. (2006). Multi-channel strategy in business-to-business markets: Prospects and problems. *Industrial Marketing Management*, *36*(1), 4–9. doi:10.1016/j.indmarman.2006.06.010

Sachan, A., & Datta, S. (2005). Review of supply chain management and logistics research. *International Journal of Physical Distribution & Logistics Management*, *35*(9), 664–705. doi:10.1108/09600030510632032

Salvador, F., Rungtusanatham, M., & Forza, C. (2004). Supply-chain configurations for mass customization. *Production Planning and Control*, *15*(4), 381–397. doi:10.1080/0953728042000238818

Santoso, T., Ahmed, S., Goetschalckx, M., & Shapiro, A. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, *167*(1), 96–115. doi:10.1016/j.ejor.2004.01.046

Sarac, A., Absi, N., & Peres, S. D. (2010). A literature review on the impact of RFID technologies on supply chain management. *International Journal of Production Economics*, *128*, 77–95. doi:10.1016/j.ijpe.2010.07.039

Shapiro, J. F. (2004). Challenges of strategic supply chain planning and modeling. *Computers & Chemical Engineering*, *28*, 855–861. doi:10.1016/j.compchemeng.2003.09.013

Sheffi, Y. (2001). Supply chain management under the threat of international terrorism. *The International Journal of Logistics Management*, *12*(2), 1–11. doi:10.1108/09574090110806262

Shen, Z. M. (2007). Integrated supply chain design models: A survey and future research directions. *Journal of Industrial and Management Optimization*, *3*(1), 1–27. doi:10.3934/jimo.2007.3.1

Sheu, J. B. (2003). Locating manufacturing and distribution centers: An integrated supply-chain based spatial interaction approach. *Transportation Research Part E, Logistics and Transportation Review*, *39*, 381–397. doi:10.1016/S1366-5545(03)00018-8

Shin, T.-H., Chin, S., Yoon, S. W., & Kwon, S. W. (2011). A service-oriented integrated information framework for RFID/WSN-based intelligent construction supply chain management. *Automation in Construction*, *20*(6)..doi:10.1016/j.autcon.2010.12.002

Slack, N., Chambers, S., & Johnston, R. (2001). *Operations management* (3rd ed.). Harlow, UK: Pearson Education Limited.

Snyder, L. V., Scaparra, P. M., Daskin, M. S., & Church, R. L. (2005). *Planning for disruptions in supply chain networks: Tutorials in operations research*. New Orleans, LA: Informs.

Sodhi, M. S., & Tang, C. S. (2009). Modeling supply-chain planning under demand uncertainty using stochastic programming: A survey motivated by asset–liability management. *International Journal of Production Economics*, *121*, 728–738. doi:10.1016/j.ijpe.2009.02.009

Srivastava, S. K. (2007). Green supply-chain management: A state-of-the-art literature review. *International Journal of Management Reviews*, *9*(1), 53–80. doi:10.1111/j.1468-2370.2007.00202.x

Stadtler, H. (2009). A framework for collaborative planning and state-of-the-art. *OR-Spektrum*, *31*, 5–30. doi:10.1007/s00291-007-0104-5

Swaminathan, J. M., Smith, S. F., & Sadeh, N. M. (1998). Modeling supply chain dynamics: A multi-agent approach. *Decision Sciences*, *29*(3), 607–633. doi:10.1111/j.1540-5915.1998.tb01356.x

Tarantilis, C. D. (2008). Topics in real-time supply chain management. *Computers & Operations Research*, *35*(11), 3393–3396. doi:10.1016/j.cor.2007.01.028

Toktas-Palut, P., & Ulengin, F. (2011). Coordination in a two-stage capacitated supply chain with multiple suppliers. *European Journal of Operational Research*, *212*(1)..doi:10.1016/j.ejor.2011.01.018

Towill, D. R. (1997). Forridge - Principles of good material flow. *Production Planning and Control*, *8*(7), 622–632. doi:10.1080/095372897234722

Van der Vorst, J., & Beulens, A. (2002). Identifying sources of uncertainty to generate supply chain redesign strategies. *International Journal of Physical Distribution & Logistics Management*, *32*(6), 409–430. doi:10.1108/09600030210437951

Van Maanen, J. (1995). Fear and loathing in organization studies. *Organization Studies*, *6*(6), 687–692. doi:10.1287/orsc.6.6.687

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, *22*(2), 195–291. doi:10.1108/01443570210414329

Voudouris, V. T. (1996). Mathematical programming techniques to debottleneck the supply chain of fine chemical industries. *Computers & Chemical Engineering*, *20*, 1269–S1274. doi:10.1016/0098-1354(96)00219-0

Wacker, J. G. (1998). A definition of theory: Research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, *16*, 361–385. doi:10.1016/S0272-6963(98)00019-9

Wagner, S. M., & Neshat, N. (2010). Assessing the vulnerability of supply chains using graph theory. *International Journal of Production Economics*, *126*, 121–129. doi:10.1016/j.ijpe.2009.10.007

Wanke, P. F., & Zinn, W. (2004). Strategic logistics decision making. *International Journal of Physical Distribution & Logistics Management*, *34*(6), 466–478. doi:10.1108/09600030410548532

Warburton, R. D. H., Hodgson, J. P. E., & Kim, Y. K. (2004). *An analytical investigation of the bullwhip effect.* National Textile Annual Report.

Williams, J. F. (1981). Heuristic techniques for simultaneous scheduling of production and distribution in multi-echelon structures: Theory and empirical comparisons. *Management Science*, *27*(3), 336–352. doi:10.1287/mnsc.27.3.336

Williams, J. F. (1983). A hybrid algorithm for simultaneous scheduling of production and distribution in multi-echelon structures. *Management Science*, *29*(1), 77–92. doi:10.1287/mnsc.29.1.77

Woodall, W. H., & Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, *31*(4), 376–386.

Xia, D., & Chen, B. (2011). A comprehensive decision-making model for risk management of supply chain. *Expert Systems with Applications*, *38*, 4957–4966. doi:10.1016/j.eswa.2010.09.156

This work was previously published in Supply Chain Sustainability and Raw Material Management: Concepts and Processes, edited by Reza Zanjirani Farahani, Shabnam Rezapour and Laleh Kardar, pp. 82-100, copyright 2012 by Business Science Reference (an imprint of IGI Global).

Chapter 102
Green Computing as an Ecological Aid in Industry

Oliver Avram
Ecole Polytechnique Fédérale de Lausanne, Switzerland

Ian Stroud
Ecole Polytechnique Fédérale de Lausanne, Switzerland

Paul Xirouchakis
Ecole Polytechnique Fédérale de Lausanne, Switzerland

ABSTRACT

This chapter concerns the use of computing for ecological evaluation in the manufacturing industry. Here, ecological evaluation means identification and quantisation of various manufacturing process characteristics from the point of view of the environment. Manufacturing is a complex process with many different interactions between the parameters controlling the manufacturing machine tools. In the past, manufacturing planners and operators have set these parameters without understanding the consequences, leading to waste of energy, cutting fluid, and so on. This chapter presents a computer tool for evaluating and quantifying the effects of different manufacturing choices using chosen criteria. The tool was implemented as part of the work for a European project. It is based on an extensive analysis of machine tools to provide a way of handling the complexities of understanding the use phase of products.

INTRODUCTION AND BACKGROUND

The aim of this chapter is to explain the use and practice of one aspect of green computing in the manufacturing sector. The contents of the chapter concerns work done for a doctoral dissertation (Avram, 2010) performed as part of a European project on the next generation of machine tools (the NEXT project). The work is important for manufacturers to evaluate the consequence of different manufacturing strategies and to quantify the effects of these on the environment. Manufacturing is a very complex task with many interactions making it difficult or impossible to understand the consequences of manufacturing decisions.

The major environmental aspects of machine tool use come during the use phase of the machine tool. At the same time, existing evaluation tools

DOI: 10.4018/978-1-4666-1945-6.ch102

are not sensitive enough to cope with the multiple manufacturing variants. Some work has been reported in the literature (Munoz & Sheng, 1995; Dornkudwar, et al., 1998; Srinivasan & Sheng, 1999a, 1999b; Akbari, et al., 2001; Dahmus & Gutowski, 2004; Jayal & Balaji, 2007; Narita, et al., 2006; Zhigang, et al., 2008, etc.) considering green machining issues; however, in most cases the focus has been limited at the process and process planning level.

Munoz and Sheng developed a model with the three dimensions of material, energy, and time as a basis. The material inputs are the primary material and secondary catalyst material, with output of primary material, waste primary material, and waste catalyst. For the energy dimension, Munoz and Sheng identified thermal, electrical, and chemical as input energy with thermal, chemical, and radiative energy as possible outputs. For the time, there is the initiation time as input and the processing time as output. They found that the geometry of the part, the workpiece material and cutting fluid were the dominant factors affecting process energy.

Domkundwar et al. used features as a basis for analysis for environmental impact assessment. The features were treated on a global basis with additional process or process parameter constraints.

Srinivasan and Sheng worked on quantifying health hazards in wet and dry machining of a simple part with one drilling operation and two end-milling operations. No account was taken of solid particles released in completely dry machining.

Akbari et al. worked with energy consumption in the machining process. An interesting element introduced in this work was the notion of fixed energy consumption versus dynamic energy consumption. This energy is needed for production but is not connected with the production itself. In addition, there is the energy needed for the production itself which can be reduced by improving manufacturing strategies.

Dahmus and Gutowski also worked with energy analysis of the machining process. They categorised the energy consumed into three main types: "Constant start-up operations," "Run-time operations," and "Material removal operations." Based on a set of usage scenarios, which they defined, they compared the energy needs for machining both steel and aluminium on modern and conventional machines.

Jayal, Balaji et al. compared drilling of aluminium with both dry and wet cutting conditions. They concluded that flood application for drilling was needed for high accuracy.

Narita et al. developed a method for estimating the environmental burden of machining in terms of contribution to the global warming potential. They converted different elements, like the electrical consumption, the coolant, lubricants and so on, into equivalent CO_2 emissions and used the total as a measure of the environmental effects.

Zhi et al. investigated the environmental impacts of cutting fluids for cutting processes. They also studied some design aspects for machine tools and thermal deformations of both workpieces and machine tools during cutting.

Informatics is needed as a tool for industry, this can be thought of as 'green applied computing.' The work mentioned above needs to model manufacturing processes explicitly or implicitly and to calculate and present a multitude of diverse factors. This is an application of computing for improving environmental awareness, hence the term "green applied computing." Industry is doing a great deal to improve the environmental impact of manufacturing and computing tools are needed to understand and quantify the effects of manufacturing options. In addition, by exposing students to the topic, future engineers are made aware of their social responsibilities and how they can be managed. This research, on measurement and quantification, is important and needs to be extended so as to gain a better picture of environmental loading. The chapter will describe how green computing can be applied in an industrial sector, which is an important user of resources.

Machine Tool Systems

Machine tool systems are described as "mother machines" by Suh et al (2008, p. 3). in the sense that it is a machine that makes machines. There are, now, a great many types of machine tool, milling machines, turning machines, grinding machines, EDM machines as well as complex machines capable of performing multiple functions. Machine tools are complex systems with many elements to handle. In addition, the application area of machine tools involves many other variables concerning physical properties of materials, thermal and chemical effects. This chapter focuses on the use of machine tools for so-called "green machining." There are, too, many different manufacturing processes and machine tool types. The main manufacturing processes are milling and turning, but there are other processes, which are used as well, such as EDM, additive processes, laser cutting, stamping, pressing or welding. For the purposes of the research, the milling process was analysed in order to establish the principles for eco-evaluation. However, the analysis method is extensible.

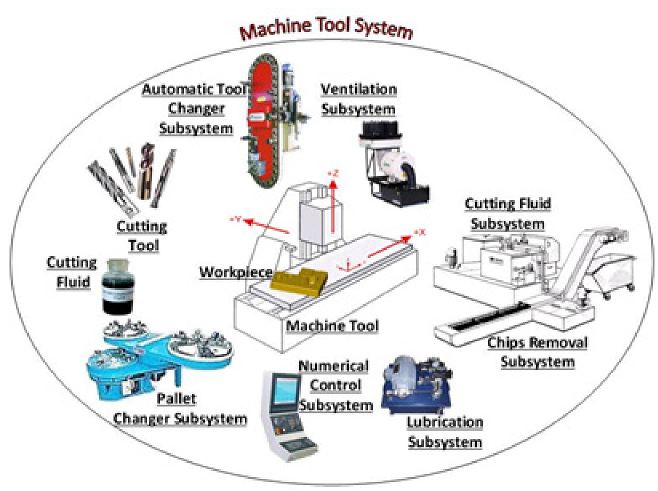
Figure 1 shows a stylised milling machine tool to illustrate the elements. There is a spindle, which holds a rotating cutting tool. The tool is moved in the Y- and Z-directions relative to the workpiece to be machined, which can be moved along X axis. Depending on the architecture of the machine the tool and the workpiece can be moved separetely or in combination.

In addition to the direct cutting processes, there are a number of subsystems for chip removal, lubrication of various components, pumps to supply cutting fluids or fans for cooling, which needed to be taken into account.

Life Cycle Costs of Machine Tools

To understand the ecological effects of any product, including machine tools, it is necessary to consider the life cycle of the product and divide it up into phases. In the design phase, it is possible to consider a number of product variants and optimise these according to different criteria. During manufacturing, the environmental effects of different manufacturing options can be evaluated for the part. The major environmental impact of many manufactured products, including machine tools, comes during the use phase. Finally, there is an End-Of-Life (EOL) phase when there are

Figure 1. The main subsystems of a machine tool system (MTS)



reuse, recycling, disposal options. The lifecycle cost of a machine tool is the sum of all these multiple effects.

As stated above, the most important environmental effects come during the use phase of the machine tool. This is also the most complicated phase, with an interaction between the ecological effects of the part and the environmental effects from the machine tool due to the way it is used for the manufacturing process. This is illustrated in Figure 2.

Use processes are performed by executing functions of the MTS, but the environmental impact does not result just from deliberated functions. Rather, they result from the behavior of the MTS subsystem responsible for that function. Therefore, it is important to describe the MTS's behavior not just its functions.

The evaluation of the use phase of the product can be simulated by determining settings of different activities within its utilisation such that the environmental impact can be reduced (i.e. finding the optimum combination of cutting parameters and their optimal sequencing among several possible combinations for the same process). The overall goal would be to propose the best-practice solution for each sub-function in order to improve the performance of the overall system.

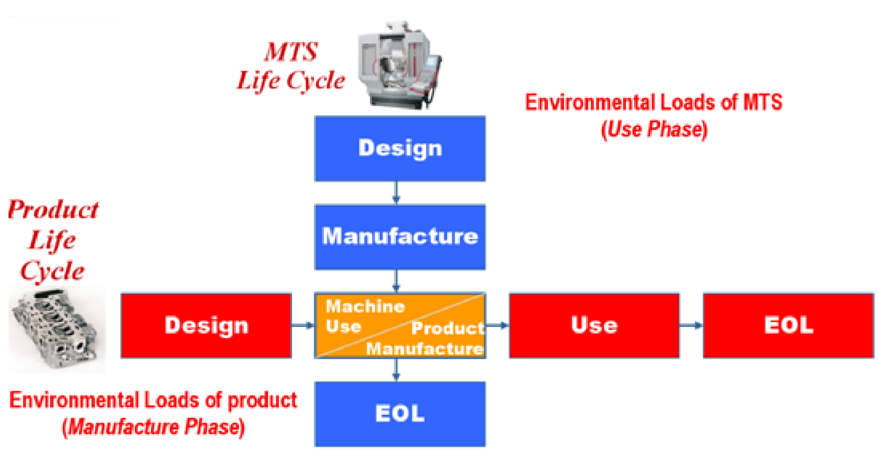
There are different environmental aspects, which need to be estimated. One important aspect is energy use. Another aspect is the resource use of potential noxious elements as waste, like cutting fluids, or in fumes, particles, etc. released into the air during machining.

EVALUATION OF THE USE PHASE

The methodology developed (Avram, et al., 2011) was based on extensive analysis and measurements of several machines. The experimental data was recorded in a structured relational database. The experiments were carried out to machine different manufacturing features under different conditions and on different machine tools. With the result data, it has been possible to evaluate and quantify the effects of manufacturing decisions. An experimental software tool was developed to show the research results.

The software tool, GREEM (Global Reasoning for Evaluation of Eco-Machining), worked on two levels. On the first level the user was asked to identify the manufacturing features to be machined, to identify the machining conditions and finally to identify the environmental aspects of interest. This data was then matched to experimental results from the database in order

Figure 2. Life cycle crossing point of a machine tool system



to retrieve relevant measurements. The user is also expected to provide weights for the different environmental aspects as a way of adapting the results to local conditions. For example, if a company uses renewable energy but has waste disposal concerns, then lubrication fluid might be set to have a higher priority than energy use. In this way the tool can be tailored to specific company needs. The results of the analysis are presented both in graphical form and as a textual form, which can be included in a manufacturing report.

While the first level is mainly concerned with aspects related to machining a specific part by an MTS, the second level assesses the machining capabilities of the MTS in terms of component specifications and activities, during its useful life when various shapes and materials have to be machined. Based on the specifications given by the MTS manufacturer for its various components and an energy consumption predictor module (Avram & Xirouchakis, 2011), the goal at the second level is to find the best performing machine tools system to complete a specific machining job.

In addition to the theoretical verification, this model has also been verified using experimental measurement of known machine performance under known conditions. The machines were run both with and without actual cutting so that the effects of the machine tool itself and the effects due to cutting could be measured. The experiments were run under different cutting conditions and with the machine tool moving at different speeds. As well as the main cutting, energy is also used by subsidiary systems and these, too, were measured for quantification. Using the machine tool model, it is possible, also, to estimate the future energy use of machines under development and to optimize the design at an early stage. For industrial users, energy minimization is important both in economic terms as well as for production optimisation.

GREEM Level 1

The first level consists of the assessment of the performance of a specific machining process with respect to a set of criteria. The machining process is defined at an elementary level, that is to say, with a unique tool class, a unique tool path, a set of cutting conditions and a cooling/lubrication alternative. The assessment can be extended at part level by gathering together under one score the performances of the machining process components carried out in order to complete the part.

The steps to be carried out for the analysis are:

- Input/Select the machine tool system
- Input/Select the machining features
- Input/Select the machining alternatives
- Input/Select the measured criteria
- Retrieve the experimental data

On this level, GREEM evaluates the effects of machining with particular options according to selected criteria. The machining description is given as a set of machining features. This method of giving machining instructions is termed "micro process planning" and derives from a new standard for machine tool control, ISO 14649 (2001). The standard allows direct use of Computer-Aided Design (CAD) data, another applied computing area, in machine tool control. The 'features' are elements like holes, pockets, face milling.

The user is first asked to define the workpiece size and material for mass property calculations for example. The user then defines the manufacturing operations in terms of manufacturing features, but only giving their gross geometric characteristics rather than the exact geometry. The user is also asked to define tool characteristics for more precise knowledge of the cutting conditions. The feature types identified for manufacturing are shown in Table 1.

An alternative, as a component of the machining strategy, refers to the choice between several possibilities all intended to accomplish a specific

Table 1. Feature types for manufacturing

Counterbore	Pattern (Rectangular)	Slot (full open)
Countersink	Spherical cap	Slot (loop)
Planar facing	Step	Slot (closed round
Hole (Round)	Thread	Slot (closed square)
Hole (Tapered)	Chamfer	Slot (closed Wood-ruff)
Pocket (Closed)	Fillet	Slot (open round)
Pocket (Open)	Profile	Slot (open square)
Pattern (Circular)	Region	Slot (open Wood-ruff)

goal. Even though the focus of this research is on the alternatives for cooling/lubrication of the cutting area, the developed methodology is also able to accommodate other alternatives (i.e. HSM—High Speed Machining). The user is also offered a choice of criteria from which to evaluate the machining plan.

Once the machining plan, the alternatives, and the criteria have been selected, it is possible to evaluate the effects of machining a part. The input data are used to select measured data from a database of experimental results. The user then has to set weights for the different criteria indicative of their relative importance. The criteria belong to the three different categories: economical, technical, and environmental. The user can either give the relative importance of the criteria in a pairwise comparison, or set their relative importance on a global scale, depending on user preference.

The software calculates the best options for machining so that users can evaluate the consequences of their machining choices. Figure 3 shows the graphical presentation of the results which can also be presented for inclusion in a machining report. On the top left, there is a presentation of the global performances based on the weights defined by the user. This shows that the best global performance for manufacturing the features chosen is MQL, or minimum quantity lubrication. MQL means that only a small amount of lubrication is delivered to the contact point between the tool, often as a mist, and the workpiece instead of 'flooding' the working area with a continuous stream of lubrication. On the top right there is a presentation of the contribution factors for machining the round hole, with workpiece quality, force and lubrication evaluations. These were the criteria chosen by the user on which to evaluate the manufacturing steps. The chart shows that for the flood option, with a large amount of cutting lubrication, the main negative environmental effect is from the lubrication itself, with a small amount of force needed for cutting but giving the best workpiece quality. For the minimum quantity lubrication there is a small contribution from the lubrication, small force needed and good workpiece quality. For dry machining, more force is needed and the workpiece quality is poorest. On the bottom left of the figure, the equivalent effects for the planar facing operation are shown. This shows that the major negative effect for flood machining is again the lubricant and the workpiece quality is not too bad. For the MQL option, the main effect is the force needed. For dry machining, much more force is needed and the workpiece quality is lower than with MQL.

The computing tools for the first level of GREEM involve mainly database tools with user-friendly presentation tools. The computer tools are important to make sense of the wealth of information surrounding the manufacturing process.

GREEM Level 2

In addition, similarly to the first level, it is necessary for the user to set priorities. The second level has to address the performance of the machining system as a whole and not just certain aspects. The software makes it possible to compare different machine tool systems with respect to various criteria, e.g. productivity, precision, cost, safety, flexibility/adaptability, and energy. The user has to emphasise the relevant importance of these criteria by using the priority settings, which use

Figure 3. Graphical presentation

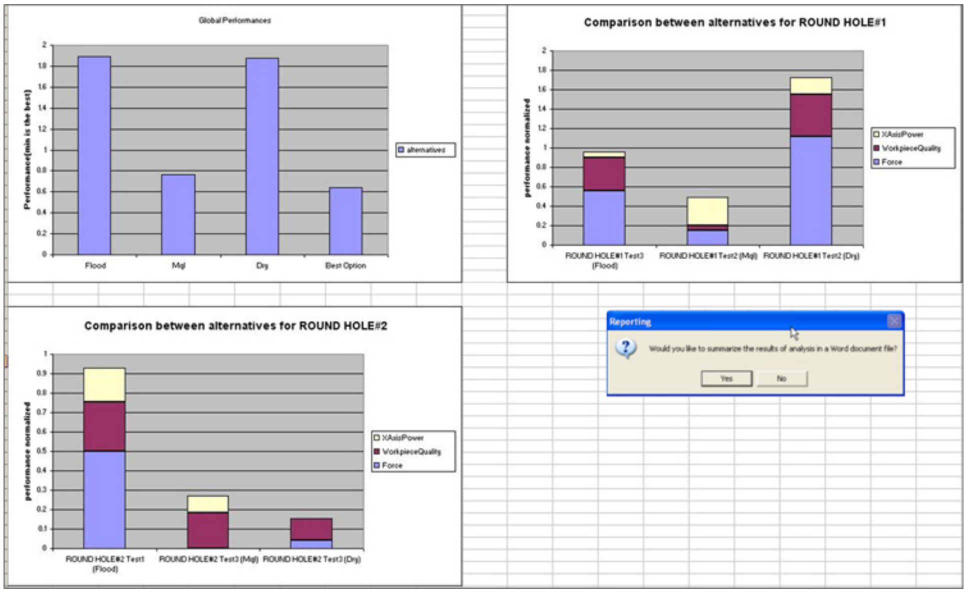
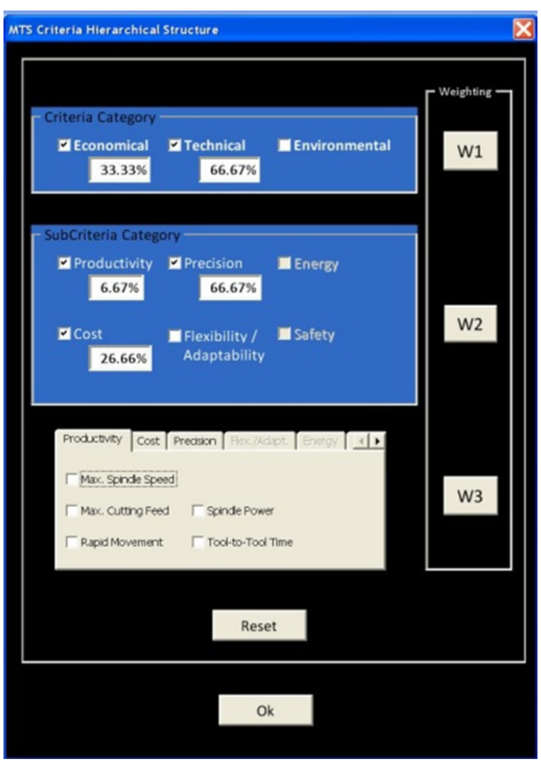


Figure 4. Priority setting

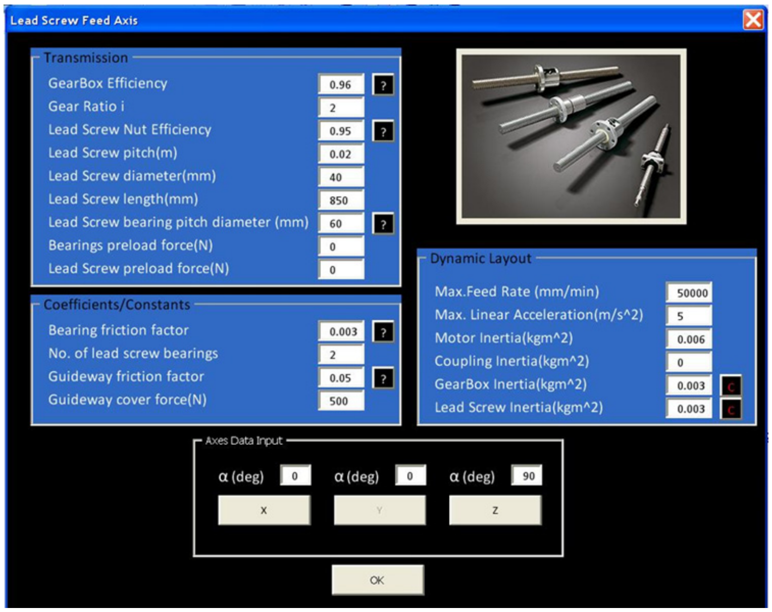


the same type of pair wise comparison technique as described earlier. Figure 4 shows the priority setting.

Energy use calculation is complicated and there are effects related to machining strategies. For the research done, a prismatic part was used. In order to calculate the expected energy use for machining a test part, it is necessary to have a model of the machine tool. Such a model has to include the masses of the machine tool elements, spindle and feed axes characteristics, kinematic structure, friction coefficients and so on. Figure 5 shows an example of input data required to estimate the energy used by one element, in this case the feed axis of the machine tool.

In order to calculate the movements of the machine it is also necessary to have the control data. For the purposes of the research, an APT (Automatically Programmed Tools) control data file was used. The APT file contains explicit movement data, which can be used to calculate the power requirements with respect to the chosen tool path. The same file was also used to mill the part while measuring the real power consumption so that the model could be verified.

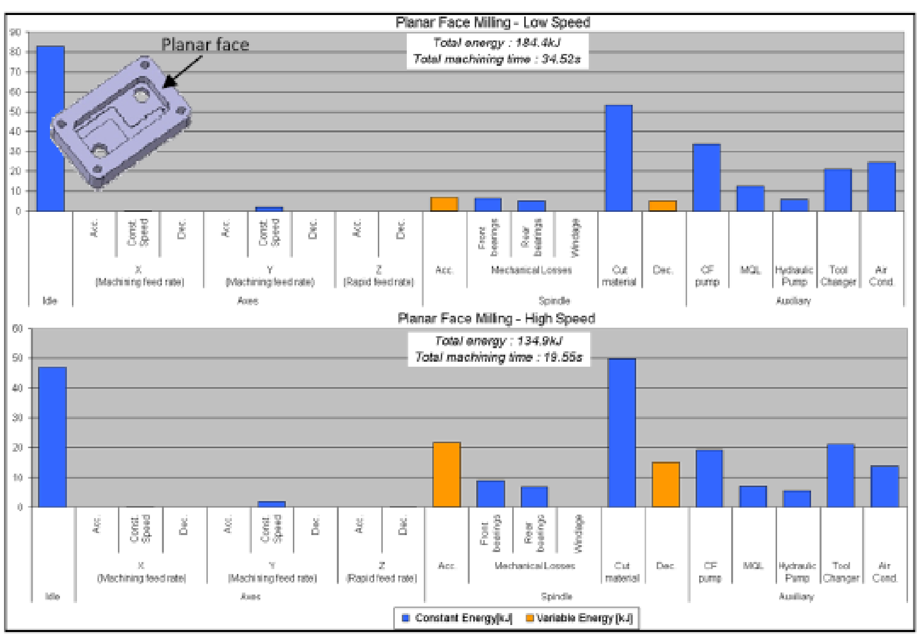
Figure 5. Feed-axis input for GREEM level 2



As with level 1 of GREEM, the analysis is divided into machining features. Figure 6 shows some results of the milling energy consumption for the planar face feature of the modelled part, machined under wet conditions and by removing the same amount of material at two different cutting speeds.

The scale on the left shows the number of kilojoules of energy consumed for different operations. On the bottom are the different elements

Figure 6. Machine tool system consumed energy (kJ) for planar face milling (low speed vs. high speed)



that were measured in the experiment. First there is the energy consumed while the machine is idling, that is, not cutting. Next, for each of the axes of the machine tool (which is a three-axis machine tool in this case), there is the energy measured while the axis was accelerating, moving at constant speed and decelerating. Then there is the contribution to energy consumption by the spindle. The acceleration and deceleration contributions are variable amounts of energy. Also shown is the energy loss due to mechanical effects of the spindle (front bearings, rear bearings, and windage). The energy consumed in actual cutting of material is also shown. The last five columns in the figure show the energy losses due to auxiliary equipment around the machine tool.

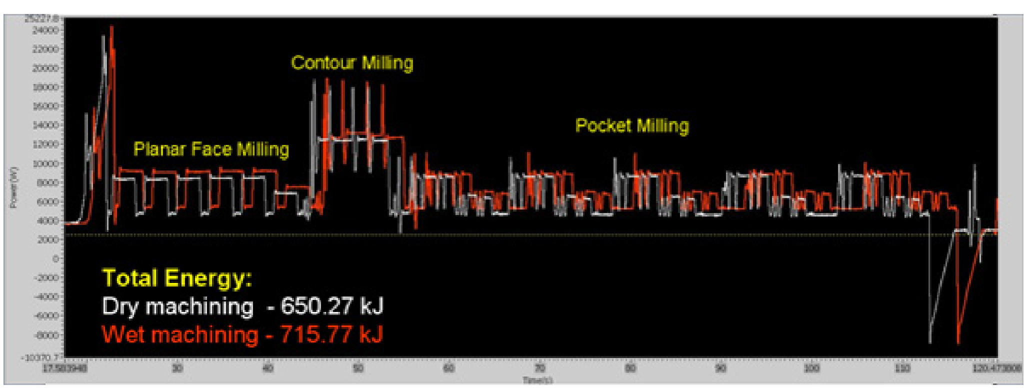
This kind of measurement provides a basis for accurate modelling of the machine tool and manufacturing processes. With this data, it is possible to help the user optimise machining parameters. It is also needed for accurate machine tool modelling for advanced machine tool control, to let the machine tool itself optimise its behaviour, which is a current research topic being actively worked on in many places.

The auxiliary equipment concerns other energy consuming subsystems on the machine tool, such as chip removal systems and pumps for providing cutting fluids. The energy consumed by the pump for the delivery of the cutting fluid (CF pump in the Figure 6) is notable because the pump cannot be switched off unless a part is to be machined 'dry.' The power drawn by the pump is highly influenced by the fact that a specific pressure should be maintained and frictions in pipes and fittings as well as elevation differences should be overcome across the length of the process. A more energy efficient alternative is the Minimum Quantity Lubrication (MQL).

Avram (2010) provides an extensive presentation of opportunities for improvement of the environmental impact of machine tools during their use phase. Due to space limitations, only a single example is illustrated here. Figure 7 shows the power profiles acquired during the machining of three milling features by employing identical cutting conditions but different cooling/lubrication strategies. Another interesting conclusion was that high-speed machining, a process where the tool spindle is rotated faster and the axis movement speed increased, can be more economical than conventional machining. Although high-speed machining needs more power, this is applied for shorter periods resulting in a net saving of energy.

Figure 7 shows the power consumption during milling of the example prismatic part used for the experiments. The scale on the left is the power consumption. Note that there is some negative consumption where power is generated in certain phases of the process as the motor becomes a generator under certain conditions. The figure shows two cutting profiles superimposed, in red

Figure 7. Machine tool system energy consumption (kJ) for part milling (wet vs. dry machining)



for wet machining and in white for dry machining. The scale at the bottom shows the time, in seconds, over which the machining took place.

For the test part, there were three machining phases: a planar face milling, a contour milling and pocket milling. The planar face milling removes a layer of material from the workpiece to make it the correct height. The profile milling cuts around the outside of the part to give it the right exterior shape. The pocket-milling phase removes material from the inside of the part.

The face-milling phase consists of a series of passes to cover the whole area of the part; the peaks in the diagram correspond to the different passes. In contour milling, the tool cuts material with the side. This is not done in one pass as cutting too deeply can lead to tool deformation and even breakage. Instead, the tool is passed round the material in a series of steps. For pocket milling, the tool is sunk into the material and then moved around to clear the material. As with profile-milling, this may not be possible in a single step and so the material clearance is done in a set of layers. The diagram shows the peak distributions of these different steps with different axes moving different amounts according to the operational characteristics. Here, only the total energy consumption is shown, not the individual contributions of the systems and subsystems of the machine tool. In order to get an accurate picture it is necessary to measure at high frequency, resulting in a huge amount of data to process.

For the second level of GREEM, database tools are also important as well as simulation, calculation and virtual instrumentation tools.

EXTENDING THE SCOPE OF THE SOFTWARE

The test plan for milling, which has been carried out on a number of machines, is shown as an example for comparison. The other test plans are suggestions for possible extensions to the database of experimental results. The milling tests were carried out, machining the same piece with three options: dry, MQL and flood machining, and with two types of tool: coated and uncoated. This means that six sets of experiments and six sets of tools were required for the tests for each set of machining parameters (feeds and speeds). The more experiments performed the more accurate the results. However, the experiments were performed with new tools for each set of measurements, making the experimental series expensive. The criteria measured were: power consumption of the main power-driven components of the machine, cutting forces and the quantity of cutting fluid applied.

It would certainly be possible to devise an extended set of tests to augment the existing database so that different process plans using different processes could be compared. Four cutting processes are considered here:

1. Milling
2. Turning
3. Wire EDM
4. Sink EDM

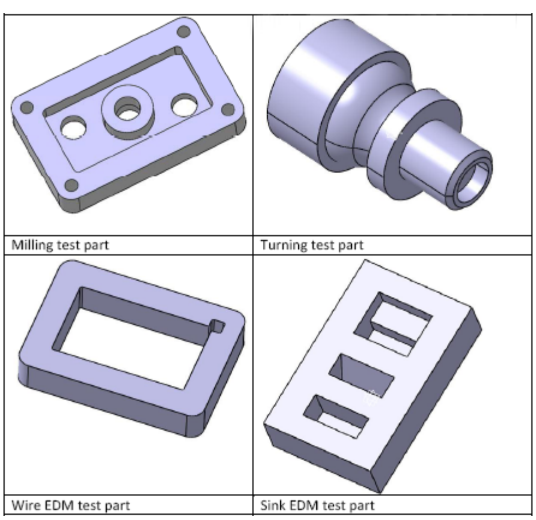
Milling is a much used process for creating general shapes using a rotating tool to remove material from a workpiece. Turning is also a conventional process in which the workpiece rotates and the tool is fixed to create rotationally symmetric parts. Wire EDM is a non-conventional process to produce accurate parts from material by cutting parts out from a block of material using a wire. Electric current is passed down the wire resulting in sparks between the wire and the workpiece. The sparks cause enough heat to be generated to vaporise, or erode, a small amount of material. As the wire is moved, it leaves a narrow cut behind it. Sink EDM is another non-conventional process used to create accurate shapes in metal for moulds and dies, for example. It works by slowly lowering a shape towards the block of material to be shape. As different parts of the shape approach the workpiece so sparks occur between the shape being lowered and the workpiece. These sparks erode away the workpiece to leave an accurately shaped cavity.

Figure 8 shows an extended set of test objects devised to provide information about a variety of manufacturing features in the different processes. These features derive from the ISO 14649 specification. For the turned part, there are profiles, chamfers and cut-ins, for example. For the wire EDM process there is a simple profile cutting feature as well as a 'pocket' feature where a complete volume of material is vaporised. For the sink EDM process, there are different types of pockets. Sink EDM is also interesting because it offers an alternative process for mould manufacture and could be compared with conventional and high-speed milling for particular examples.

RESEARCH VISION AND FUTURE RESEARCH

Figure 9 shows an image of the research vision, with an increased understanding of the interaction between different options in manufacturing. This obviously requires a far wider range of analysis than has yet been performed as well as more time and a greater investment. The milling machine tool analysis is complex enough, but the higher levels are even more complex and need more related research in order to be able to make the whole process more flexible.

Figure 8. Test parts for process extensions

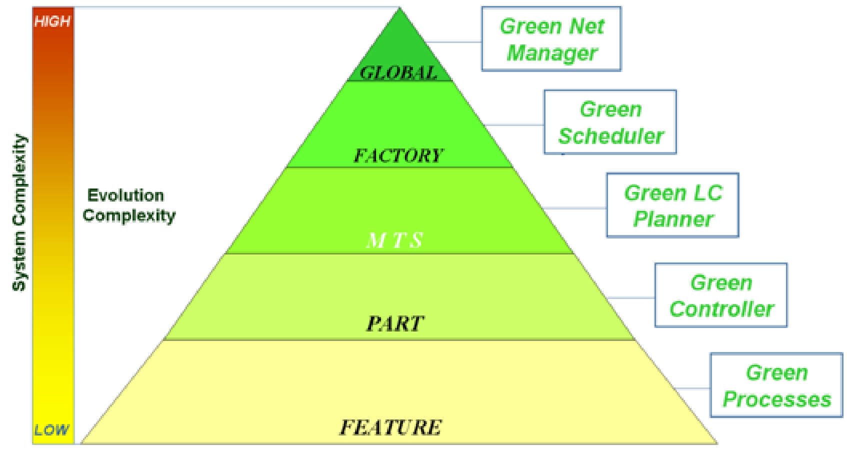


An example of the needs for change concerns the modern machine tool control standard, ISO 14649 or STEP-NC. The information provided for control by this new standard allows the machine tool to optimise its own behaviour for its architecture and the part to be machined. This is possible because of the increased computing power now available on the industrial computers used in controllers. Traditionally, movement planning had to be done off-line with neither details of the architecture of the target machine tool nor of workshop loading. An extension to this standard is to have a machine tool model to describe available manufacturing resources as well as to describe machining needs in the control file. Such an extension would allow workshop planners to make last-minute decisions about machine choice.

Another important aspect of improved level of information on the shop floor is the possibility to perform simulations and optimisation of plans. In some processes, such as EDM, the machine operators have extensive knowledge of the process and sometimes perform experiments to improve machining parameters before real production starts.

Concerning global levels, life-cycle analysis tools exist but sometimes need careful application in order to understand the results. An example of a tool, which is easier to use and can be used to compare alternatives is the EcoIndicators99 method. This method uses a scoring system to equate different environmental factors. Transport, machining, part treatments, recycling and so on are all assigned a unit score in a common basis which is then multiplied by part characteristics, such as weight, to arrive at a process score. These scores are then added together to give an idea of the relative environmental impact of the part production. The methods described above keep the different factors separate in order to understand from

Figure 9. Machine tool system hierarchy of modeling and optimization methods and tools



where the different environmental factors come, a so-called multi-criteria approach. The method presented also allows the user to understand the environmental effects in quantative terms rather than simply knowing which option may be better.

Further research is going on in other projects where advanced simulation and calculation are also being used. Data modelling based on computer science tools also plays a major part in the standardisation work.

CONCLUSION

Computing at the service of the environment has a wider application than simply improving the environmental aspects of computing itself. Green applied computing can be used to make sense of complex areas where the wealth of data is too great to be comprehensible without processing. In the manufacturing context there is a great deal of complexity already with machining but extending consideration to the factory level and the global supply chain then the complexity far exceeds the level which can be comprehended by users.

For the research presented, the main computing tools used were:

- Database tools
- Data collection and management tools
- Simulation tools
- Calculation

ACKNOWLEDGMENT

The authors would like to thank Vincent Capponi for all his work. We would also like to thank the European Union for financing the research presented during the NEXT project. Last but not least, we would like to thank the NEXT project partners for sharing their results and equipment for the experiments that were so necessary for the work.

REFERENCES

Akbari, J., Oyamada, K., & Saito, Y. (2001). LCA of machine tools with regards to their secondary effects on quality of machined parts. In *Proceedings EcoDesign 2001: Second International Symposium on Environmentally Conscious Design and Inverse Manufacturing*, (pp. 347-352). Tokyo, Japan: IEEE.

Avram, O. (2010). *Machine tool use phase: Modeling and analysis with environmental considerations*. Ph.D. Dissertation. Lusanne, Switzerland: EPFL.

Avram, O., Stroud, I., & Xirouchakis, P. (2011). A multi-criteria decision method for sustainability assessment of the use phase of machine tool systems. *International Journal of Advanced Manufacturing Technology*, *53*, 811–828. doi:10.1007/s00170-010-2873-2

Avram, O., & Xirouchakis, P. (2011). Evaluating the use phase energy requirements of a machine tool system. *Journal of Cleaner Production*, *19*, 699–711. doi:10.1016/j.jclepro.2010.10.010

Dahmus, J., & Gutowski, T. (2004). An environmental analysis of machining. In *Proceedings of the 2004 ASME International Mechanical Engineering Congress*. Anaheim, CA: ASME.

Dornkundwar, V., Krishnan, N., Bauer, D., & Sheng, P. (1998). A distributed environmental planning model for machined components. In *Proceedings 5th International Seminar on Life Cycle Engineering*, (pp. 227-239). Stockholm, Sweden: KTH. Eco-indicator-99. (2011). *Website.* Retrieved from http://www.pre.nl/content/eco-indicator-99.

ISO 14649. (2001). *TC184/SC1/WG7, ISO 14649/FDIS, data model for computerized numerical controllers*. New York, NY: International Standards Organization.

Jayal, A. D., & Balaji, A. K. (2007). On a process modeling framework for sustainable manufacturing: A machining perspective. In *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*, (pp. 301-307). ASME.

Munoz, A. A., & Sheng, P. (1995). An analytical approach for determining the environmental impact of machining processes. *Journal of Materials Processing Technology*, *53*, 736–758. doi:10.1016/0924-0136(94)01764-R

Narita, H., Kawamura, H., Norms, T., Lian-Yi, C., Fujimoto, H., & Hasebe, T. (2006). Development of prediction system for environmental burden for machine tool operation. *JSME International Journal Series C*, *49*(4), 1188–1195. doi:10.1299/jsmec.49.1188

Srinivasan, M., & Sheng, P. (1999a). Feature-based process planning for environmentally conscious machining - Part 1: Microplanning. *Robotics and Computer-integrated Manufacturing*, *15*, 257–270. doi:10.1016/S0736-5845(99)00017-4

Srinivasan, M., & Sheng, P. (1999b). Feature based process planning in environmentally conscious machining- Part 2: Macroplanning. *Robotics and Computer-integrated Manufacturing*, *15*, 271–281. doi:10.1016/S0736-5845(99)00018-6

Suh, S.-H., Kang, S.-K., Chung, D.-H., & Stroud, I. A. (2008). *Theory and design of CNC systems*. Berlin, Germany: Springer.

Zhi, G. J., Hua, Z., & Xiao, L. (2008). Study on the cutting technology based on green manufacturing. *Key Engineering Materials*, *375-376*, 158–162. doi:10.4028/www.scientific.net/KEM.375-376.158

This work was previously published in Sustainable ICTs and Management Systems for Green Computing, edited by Wen-Chen Hu and Naima Kaabouch, pp. 132-145, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 103
Improving Energy-Efficiency of Scientific Computing Clusters

Tapio Niemi
Helsinki Institute of Physics, Finland

Jukka Kommeri
Helsinki Institute of Physics, Finland

Ari-Pekka Hameri
University of Lausanne, Switzerland

ABSTRACT

The authors applied operations management principles on scheduling and allocation to scientific computing clusters to decrease energy consumption and to increase throughput. They challenged the traditional one job per one processor core scheduling method commonly used in scientific computing with parallel processing and bottleneck management. The authors tested the effect of increased parallelism by using different test applications related to high-energy physics computing. The test results showed that at best their methods both decreased energy consumption down to 40% and increased throughput up to 100%, compared to the standard one task per CPU core method. The trade-off is that processing times of individual tasks get longer, but in scientific computing, the overall throughput and energy-efficiency are often more important.

INTRODUCTION

Scientific computing clusters are widely used in many research fields, especially in experimental physics, astronomy, and bio-sciences. Computing intensive research easily deploys thousands, even hundreds of thousands, of CPUs to analyze various data sets and models. These clusters can be viewed as production resources processing jobs consisting of numerous tasks. The tasks can be processed by different resources, and finally the jobs are assembled together to be delivered back to the cluster customers. Operating such a cluster has its own cost structure related to capital invested, energy consumed, maintenance work, and facility related costs. In all, a computing cluster closely resembles an industrial production unit, thus the working hypothesis for this chapter is

DOI: 10.4018/978-1-4666-1945-6.ch103

to apply operations management principles used in manufacturing to improve computing cluster productivity and overall efficiency.

Green computing, as it is now often referred to, seeks to improve efficiency of computing centers. It is a wide topic incorporating issues like data centre locations near cheap energy sources (Brown & Reams, 2010), minimizing so called e-waste (Hanselman & Pegah, 2007), designing optimal cooling infrastructure and running the centre in an optimal way (Marwah, et al., 2009). Generally, energy and resource optimization in scientific computing has mostly focused on hardware and infrastructure issues, for example the development of more efficient hardware or optimizing cooling of computer centers, and has not focused as much on operational methods such as workload management and even less on operating systems and application software optimization for energy-efficiency.

The most well known method for comparing energy efficiency of data centers is Power Usage Effectiveness (PUE) metrics (Belacy, 2008). This is a ratio of the total facility power / IT equipment power. It indicates how much of the energy is lost in cooling, power distribution, and other infrastructures. However, it does not indicate how efficiently IT resources are operated. Limitations of PUE metrics have been recognized by the Green IT Promotion Council (2010), which proposes three additional metrics: ITEU (IT Equipment Utilization)—IT equipment usage in data center; ITEE (IT Equipment Energy Efficiency)—total rated capacity of IT equipment / total rated energy consumption of IT equipment; and GEC (Green Energy Coefficient)—Green (natural energy) energy / total energy consumption of data center. Based on these metrics it is possible to calculate Datacenter Performance Per Energy used (DPPE) metrics as follows: DPPE = (ITEU x ITEE x 1/PUE) / (1-GEC). Further, the report gives four methods for improving energy efficiency: 1) Operating the data center in an efficient way, i.e. reducing amount of hardware and increasing its utilization; 2) Installing energy efficient hardware; 3) Improving energy efficiency of non-IT infrastructure; and 4) Using renewable energy.

In this work, we focus on a typical computing problem in high-energy physics: How to process a large set of jobs efficiently. While most of the existing work on high performance computing focuses on optimizing processing time of individual computing jobs, we try to optimize energy consumption per computing job and the total processing time of the set of jobs by choosing an optimal scheduling policy.

We study this problem by using real experimental physics data, computing jobs, and dedicated computing clusters. CERN, the European Organization of Nuclear Research in Geneva, provided us with a unique possibility to experiment and test the hypothesis. The Large Hadron Collider (LHC) experiment at CERN produces about 15 petabytes of data in a year. The overall computing infrastructure comprises numerous computing clusters of alternative sizes, yet the total amount of CPUs is over 100,000 in over 140 computing centers. Efficient management of these computing resources is vital for the success of the project, which is foreseen to be active for the next 20 years.

Basically, the problem is similar to production management in any manufacturing facility. In manufacturing, this kind of optimization problem can lead to a trade-off situation: improving energy-efficiency can weaken throughput. However, within the context of the computing center our tests indicated that these two aims are not necessarily contradictory, meaning that optimizing system throughput also improves energy efficiency. Our approach is based on a conclusion that computers should run near full power or be turned off, since the fixed power consumption is around 50% of the full power of the server. This chapter is based on our earlier papers (Hameri & Niemi, 2010; Niemi, et al., 2009a, 2009b).

BACKGROUND

The theoretical background of our research comes from production optimisation research. Production engineers search ways to increase throughput of a facility with limited capacity. The principles of production, much like the law of capacity dictating that, in a steady state, all plants will release work at an average rate that is strictly less than average capacity, sets a challenge to get the best out of the facility by avoiding waste, unnecessary waiting times, shut downs and disrupting events. This means that the organization and allocation of work in production systems affects the performance of the system.

As a review of related research, we first present methods to improve efficiency at the whole data centre or at least pertaining to computing clusters. For example, Lefurgy et al. (2007) suggested a method to control peak server power consumption. The method is based on power measurement information on each computing server. Controlling peak power makes it possible to use smaller and more cost- and energy-effective power supplies. Moreover, Kappiah et al. (2005) developed a method to decrease power consumption of parallel computation in power-scalable clusters, while Conner et al. (2006) studied how energy can be saved by dynamically disabling network links in supercomputing clusters.

Another group of studies focuses on servers. Venkatachalam and Franz (2005) gave a detailed overview on techniques that can be used to reduce energy consumption of computer systems. Li et al. (2005) studied performance guaranteed control algorithms for energy management of the disk and main memory. Ge et al. (2005) studied methods based on the Dynamic Voltage Scaling technology of microprocessors and created a software framework to implement and evaluate these methods. Yuan and Nahrctedt (2002) studied the same issues in mobile devices. Essary and Amer (2008) optimized disk arm movements for saving energy, while Zhu et al. (2005) proposed a disk array energy management system. Finally, Zhang et al. (2004) gave compiler-based strategies to optimize cache energy consumption of microprocessors.

One relevant field of research for our study is scheduling methods on how the order and allocation of computing tasks to computing nodes affects the efficiency of the system. The way individual computers schedule their processes is beyond the scope of this chapter. Scheduling problems can be classified according to the following properties:

- on-line / offline
- knowledge on jobs
- knowledge on computing resources

More formally (e.g. following Pinedo [2008] or Brucker [2007]), the scheduling problem can be defined as follows: We have m machines (i.e. computing nodes in our case) and j jobs to be processed. A schedule S is an allocation of time intervals from machines for each job. The challenge is to find an optimal schedule for jobs when taking into account certain constraints. A schedule is optimal if it minimizes a given optimality criteria, for example time, cost, or usage of some resources. The optimality criteria can be defined in several ways, such as the completion time of the last job or the total completion time i.e. sum of all completion times, or in our case, energy efficiency. The overall objective can also be a (weighted) sum of the sub-objectives, which often leads to a Pareto-optimal schedule.

Scheduling is a widely studied topic but most of the work focuses on finding optimal schedules when jobs have proceeding constraints and/or strict time limitations. Optimizing total throughput or energy efficiency in high throughput computing has received less research interest. Instead, some works suggest clearly opposite approaches: For example, Koole, and Righter (2008) suggest a scheduling model in which tasks are replicated to several computers. However, the authors do not estimate how much more resources are needed when the same tasks (or at least parts of them) are computed several times. Fu et al. (2007) present a scheduling model capable to restart batch jobs. They give an efficient algorithm to solve the problem but they do not mention resource usage. Further, Da Silva et al. (2003) presented another model using replication.

Most of the research on scheduling focuses on theoretical aspects but some practical studies do exist. For example, Etsion and Tsafrir (2005) compared commercial workload management systems focusing on their scheduling systems and default settings. According to the authors, the default settings are often used by the administrators, or they are just slightly modified. Goes et al. (2005) studied scheduling of irregular I/O intensive parallel jobs. They noted that CPU load alone is not enough but all other system resources (memory, network, storage) must be taken into account in scheduling decisions.

More on the theoretical side, Prasanna and Musicus (1996) define a theoretical scheduling model in which the number of processors allocated to a task can be a continuous variable, which makes it possible to allocate all processors for one task if needed. Edmonds (2000) studied non-clairvoyant scheduling in multiprocessor environments. In his model, the jobs can have arbitrary arrival times and execution characteristics can change. Santos-Neto et al. (2004) studied scheduling in case of data-intensive data mining applications, while Wang et al. (2009) developed optimal scheduling methods in a case of identical jobs and different computers. Their on-line algorithm aims to maximize the throughput and to minimize the total load. Finally, self-learning scheduling models also exist, such as that of Shivam et al. (2006).

One group of scheduling research focuses on grid computing. Medernach (2005) studied workload in a grid-computing cluster in order to be able to compare different scheduling methods. The idea of the work was to find ways to group cluster users to characterize their usage. The scheduling was based on one job per CPU core idea. Kurowsk et al. (2008) studied two-level hierarchical grid scheduling taking into account all stakeholders of grid computing systems. The approach does not require time characteristics of jobs being known. Aziz and El-Rewini (2008) studied online scheduling algorithms based on evolutionary algorithms in the grid context. Ni et al. (2005) developed a heuristic scheduling algorithm for grid environments. The algorithm is based on the concept of changing tasks between computers. Some studies apply fuzzy logic and other novel optimization methods. For example, Cao et al. (2003) applied fuzzy time functions to a large-scale grid workflow management and Liouane et al. (2008) studied multi-objective scheduling with a fuzzy controller. Some evolutionary approaches, such as Moallem and Ludwig (2009) and Grosan et al. (2007) have also been studied. Finally, there are some studies on energy-aware scheduling. For example, Rajan and Yu (2008) and Mukherjee et al. (2009) studied this topic. Further, Bunde (2006) developed power aware scheduling methods for minimizing energy consumption and not reducing system performance by applying dynamic voltage scaling technologies.

Other related work (e.g. Mu'alem & Feitelson, 2001; Tsafrir, et al., 2005, 2007) investigates the accuracy of user time estimates and noticed that they are inaccurate. To solve the problem, Tsafrir et al. developed a model to simulate user run time estimates. Piro et al. (2009) studied a similar problem in the Grid context. Yom-Tov and Aridor (2008) presented a self-learning algorithm to estimate resource requirements of batch jobs. Moreover, Karlsson et al. (2005) introduced a method to control storage access in a data centre. The method is based on adaptive controllers and does not assume prior knowledge of the system.

Virtualization has also been studied as a novel method to improve energy efficiency of computing centers. Virtualization makes it possible to combine services from several physical machines on to one physical machine. This is helpful when the utilization of the machines is minimal and the services are running on dedicated hardware. With virtualization, it is possible to use excess resources and downsize the hardware pool. However, consolidation of jobs to fewer servers is not directly relevant to our problem since we assume that there are a large number of jobs to be processed. This is a realistic assumption in physics computing. Increased security provided by virtualization is not usually needed in scientific computing.

The allocation and management of virtualized resources is a widely studied area. Both dynamic and static solutions have been proposed. For example, Bobroff et al. (2007) developed a dynamic system that forecasts future load and in this way it can minimize the number of physical servers. Further, Viswanathan et al. (2011) studied energy efficiency of virtual servers with predefined loads. They succeeded in improving energy efficiency by scheduling predefined virtual machine loads to minimize resource contention and maximize resource usage. However, running several virtual machines on one physical computer does not come without penalties. Apparao et al. (2008) studied the effects of consolidation running several benchmarks in a virtualized environment. This study indicates that sharing physical resources and resource contention clearly reduces the performance of individual virtualized applications. Pradeep et al. (2007) carried out a performance study of two different virtualization technologies. They noticed that some overhead always exists and the size of it depends on the used virtualizing technology. Like Pradeep et al. (2007), Regola et al. (2010) evaluated different virtualization technologies and concluded that full virtualization technologies still pose heavy overheads and are not viable in high throughput computing that use I/O excessively. However, virtualization would allow scientists to use more heterogeneous environments and send their jobs to computing clusters in a form of a virtual machine.

As shown above, research on improving efficiency in computing systems focuses on hardware optimization, power management, various scheduling models etc. However, there is little research on how to apply production management principles to the management of computing resources, i.e. viewing a computing centre as a factory with limited capacity and bottleneck resources. According to the manufacturing and production management point of view, the key issues are related to the optimization of throughput and minimizing operating costs; these are also desirable objectives when computing resources are managed efficiently. Our aim is to increase the utilization rate of bottleneck components (usually I/O or CPU) and therefore increase throughput, which decreases average power consumed by a computing job.

PROBLEM AND METHODS

The main research question of our study is: How can computing centers improve throughput and energy-efficiency by scheduling jobs based on their estimated properties and load requirements of all components (CPU, memory, and network) in computing nodes. A computing system is much more flexible than a manufacturing system since, in a computer, it is possible to run several tasks in parallel, and since tasks are not dedicated to any specific resource, set-up times between different jobs are also practically non-existent. Another difference with manufacturing is that it is difficult to see what is happening inside the computer, thus all measures are based on the in- and output of the system, as actual work in progress takes place literally in a black box.

The motivation for our work comes from the CERN's LHC experiment. The Worldwide LHC Computing Grid (WLCG1) is used to analyze the data produced by the Large Hadron Collider at the CERN, and it consists of hundreds of terabytes of data storage and tens of thousands CPU cores. On this scale, even small system optimization can offer noticeable energy and cost savings and performance improvements. Since high-energy physics computing has many special characteristics, common industry practices are not always the best. The main characteristics of LHC computing are: jobs contain large sets of similar tasks, data-intensive computing, processing time of an individual task is not critical, no preceding conditions among tasks, and little intercommunication between tasks, i.e. high parallelisms.

We use the following basic terminology in this work: A task is the smallest elementary entity when work is processed. The task starts, retrieves/reads its possible input file, processes the data, and possibly writes its output file. A job is a collection of tasks. In our current work, we assume that the processing order of the tasks inside a job is irrelevant. By energy efficiency, we mean the number of similar tasks or jobs that can be processed by using the same amount of electricity while throughput means the number of similar tasks or jobs that can be processed in a time unit.

We started our study by collecting and analyzing log data for a physics-computing cluster. When analyzing the data we noticed that lead times of jobs are 15 to 50% longer than the actual CPU times. Since the cluster configuration was set to process one job per CPU core, this means that there is a bottleneck slowing down the computing process. The most obvious bottleneck is I/O access to the disk and network. When estimating memory utilization of jobs compared to their CPU utilization, the memory utilization rate was about half of the CPU utilization rate. One reason for this is an irregular memory utilization curve of physics jobs. We assume that the other reason is related to I/O waiting times. Based on this analysis we assumed that reasonable overloading and processing different kinds of tasks in parallel can improve throughput by keeping bottleneck resources busy. The following two hypotheses document this:

1. Throughput can be improved and electricity consumption reduced in data intensive High-Energy Physics (HEP) computing compared to the traditional single task per CPU core processing by multitasking, i.e. processing more than one task per CPU core in parallel.
2. The performance can further be increased by mixing heterogeneous tasks while multitasking.

The hypotheses above do not give any upper limit for the number of parallel tasks. In practice, the number is limited by physical resources, especially by memory. The practical challenge in multitasking is to find the optimum number of simultaneous tasks. We used the following criteria: 1) a task uses a minimal amount of energy, and 2) the total throughput is maximal. These goals could be contradictory, but in all our tests, we noticed that the maximal throughput also gives the minimal energy consumption per processed task. As far as the clusters are concerned, the problem can be divided into two independent steps assuming that cluster nodes are homogeneous:

1. Finding the optimal load combination for the computing node.
2. Scheduling jobs to the computing nodes in such a way that all computing nodes are as close as possible to the optimum state (i.e., Step 1).

In an important special case in which all tasks are identical, the problem becomes: How many tasks must be processed simultaneously in a computing node.

We tested the hypotheses by building a realistic test environment, and then developing, and finally testing different scheduling methods. In practice, we did this by processing a job, i.e., a large set of tasks and measuring the time and electricity consumed during the test run. As explained above, we assume having a large set of independent tasks, and we are interested in the total processing time and electricity consumption of this set. By the total processing time, we mean the total time from the submission of the first task in the set to the computing cluster to the end of processing the last task in the same set. Our aim is to minimize both the amount of electricity and time needed to process the set. Minimizing the processing time is equal to maximizing the throughput.

Our experimental system was comprised of one front-end computer running the workload management system and one computing node and 1 Gb local area network. To make sure that results are not valid for one particular computer, we used two different architectures: 1) a Dell PowerEdge SC1435 server with two 4-core ADM Opteron 2376 2.3GHz and 32 GB of memory, and 2) a Dell PowerEdge R410 server with two Intel Xeon E5520 quad core 2.27 GHz processors and 16 GB DDR3 memory. The first server was used in our first test and the second server in tests two and three. We used a Sun Grid Engine (SGE) (Sun, 2008) as a workload management system, as it is also commonly used in grid computing clusters. It has various features to control scheduling. The scheduling is based on the load of the computing nodes and resource requirements of the jobs. There are also various other batch scheduling systems such as Torque[1], OpenPBS[2], LSF[3], and Condor[4]. These systems have different features but their basic functionality is very similar.

Electricity consumption of the computing nodes was measured with the Watts Up Pro electricity meter. The accuracy of the meter is around +-1%. The operating system used with Xeon and Opteron was Rocks 5.0 with kernel version 2.6.18.

We used five different test applications. The first three were implemented to test usage of only one resource and the fourth and fifth were real physics applications:

- The I/O test application writes and reads 300 MB files multiple times. After generating a file, the content is copied to another file 20 times. Each time the file is a bit different (small shift in numerical values) to avoid buffering.
- The CPU test application contains a long loop calculating floating point multiplication. A reminder of the index variable is also used to make compiler optimization harder.
- The memory test application reserves memory (200 MB) and fills it with numbers. After that, it reads a part of the memory and writes it to another part. This is done multiple times.
- The physics data analysis application is based on the CMSSW framework. Input data for the test is from the CRAFT (CMS Running At Four Tesla) experiment that used cosmic ray data recorded with CMS detector at the LHC during 2008 (Acosta & Camporesi, 2008). This detector was used much like current LHC experiments and the data was very close to the final data analysis. The analysis software reads the input file (94 MB or 360 MB), performs the analysis, and writes a small summary file on the local disk. Input data is stored in ROOT data container files (Antcheva, et al., 2009). Physics data is stored in binary format to save space and one root file usually contains hundreds of events captured by the detectors. The disk I/O during the application processing is shown in Figure 1 and the memory usage in Figure 2.
- The CPU intensive physics application is also based on the CMSSW framework (Fabozzi, et al., 2008) including event generation with Pythia6 (Sjostrand, et al., 2006) and full detector simulation with Geant4 (Agostinelli, et al., 2003). The Pythia program is a standard tool for the generation of high-energy collisions using Monte Carlo methods.

Each of these test sets was run several times to get stable results. The input data were different for each run to eliminate the effect of the disk cache. Additionally, the disk cache was cleared

Figure 1. I/O usage of a single physics analysis job

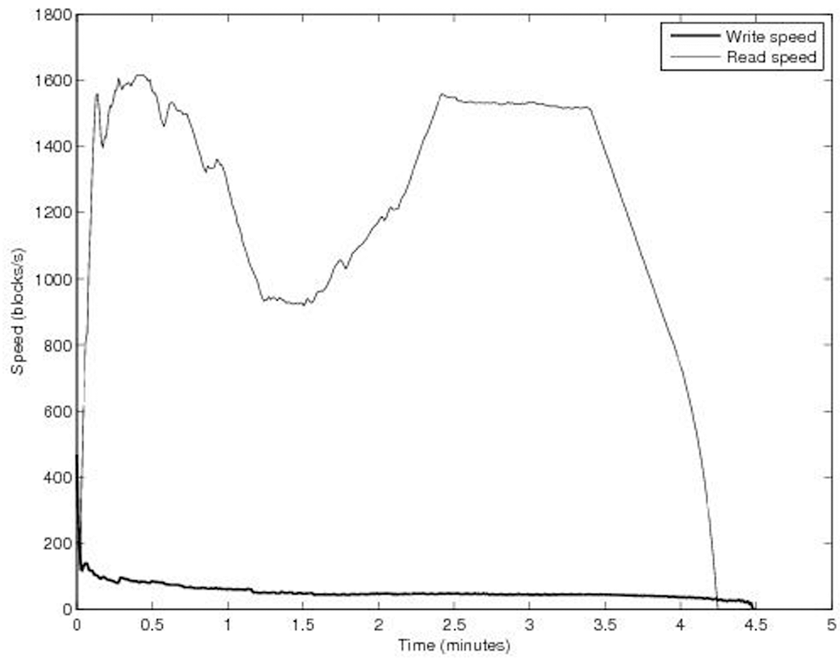
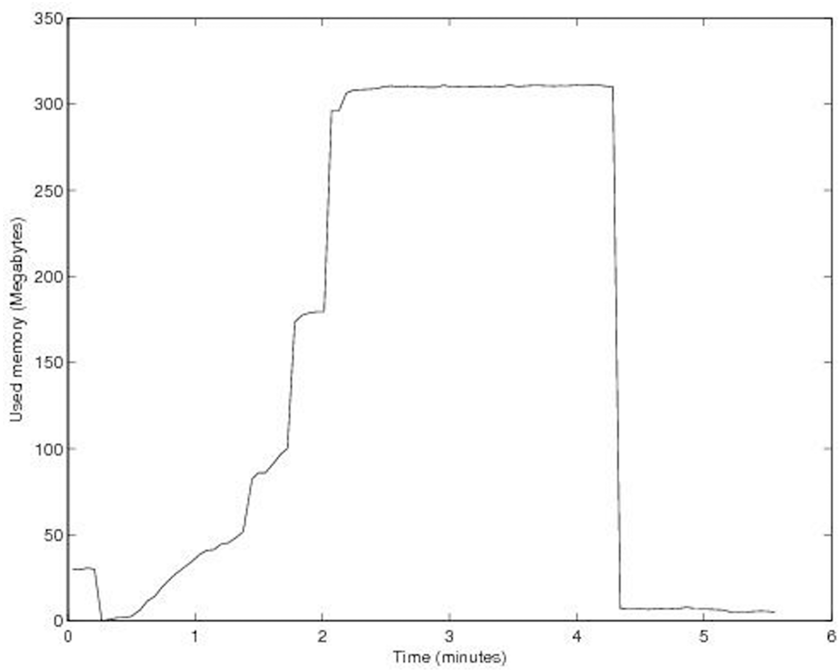


Figure 2. Memory usage of a single physics analysis job



between the test sets to provide a similar environment for all the test sets. Since differences in averages among the different test sets were relatively high and standard deviation inside the sets low, this test setting gave us large enough sample size to look for statistically significant results.

TESTS AND RESULTS

We tested our hypotheses by using simplified test applications and high-energy physics applications. In the first test we compare the performance of different applications when changing the amount of parallelism. The second test shows how the performance of the I/O application varies when it is processed with a different amount of parallel tasks. The final test demonstrates that mixing different tasks (I/O and CPU oriented) can further increase performance.

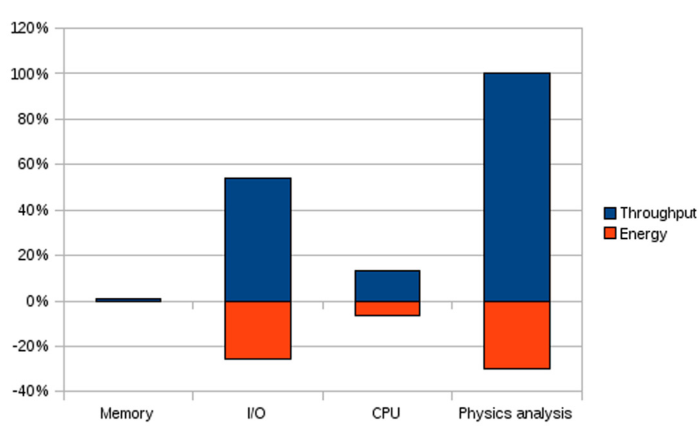
In the first test, running 3-4 tasks per CPU core yielded the best results, but the test also showed that benefit of multitasking heavily depends on the application used. If an application does almost only memory operations such as our memory test application, the benefit is almost zero. Luckily, few real applications are like this. For computing oriented CPU applications improvements are quite clear. Our first test case produced 13% more throughput, while 7% less energy was consumed. Much larger improvements are related to I/O, by processing our I/O test application using 4 tasks per CPU core setting instead of 1 task per core, improved throughput over 50% and decreased energy consumption around 25%. However, the biggest improvement gave our real physics analysis application 100% more throughput while 30% less energy was consumed. The test results are shown in Table 1 and percentile changes when comparing the best case to one task per CPU core case, are shown in Figure 3. These results validate Hypothesis 1.

The reason for the improvements is not straightforward. Intuitively one could assume that especially overloading I/O could decrease throughput.

Table 1. Test results of the first test

Test type	Jobs/core	Jobs/hour	Wh/job
Memory	1	90	2.9
	3	91	2.9
I/O	1	320	0.79
	4	493	0.59
CPU	1	407	0.6
	3	461	0.56
Physics analysis	1	188	1.1
	3	376	0.77

Figure 3. Improvements of running 3-4 tasks in parallel compared to 1 task/core



However, according to our tests, it seems to be that overloading I/O improves performance. We assume two possible reasons: 1) Overloading keeps the bottleneck busy and eliminates the total I/O waiting time in the system, and/or 2) overloading improves the efficiency of disk caches since it is more likely that some other task has already fetched the required data.

Figure 4 shows the variation of throughput and energy-consumption when processing 1 to 50 I/O tasks in a 8-core server. The difference is calculated comparing the result to 1 task per core, i.e. 8 tasks per server, setting. Each measurement was taken twice and the test application was standard physics data analysis. The figure shows that after reaching 1 task per core situation, the performance and energy-efficiency still improve almost linearly until the physical memory is full.

Our final test was related to mixing different applications in parallel processing. Since the physics analysis seemed to give the biggest benefit on multitasking, we tested the effect of mixing I/O and CPU intensive tasks by using two different physics applications: an I/O intensive data analysis and a CPU intensive simulation. We compared first processing CPU-oriented tasks and then I/O-oriented physics tasks to processing the same jobs mixed together. Figure 5 illustrates the improvements in energy consumption and throughput in these tests. The results indicated that mixing tasks using different resources clearly improved both throughput and energy-efficiency. This validates Hypothesis 2.

DISCUSSION

In scientific computing clusters, the single job (task) per core-scheduling method is mostly used. This simply means that each CPU core can run a maximum of one job. Usually the jobs are distributed equally to all computing nodes of the cluster. Instead of this common practice, we tested variations of different scheduling methods based on the idea that it is necessary to fully load computing nodes. Our tests showed that running multiple tasks simultaneously, or mixing tasks that utilize different resources, can decrease energy usage per computing task and improve throughput of the computing node when running High-Energy Physics (HEP) applications. The trade-off is that processing times of individual tasks are longer but in cases, such as HEP computing, in which the tasks are not time critical, usually only the total throughput is important. We assume that the reason for the improvement is that some component

Figure 4. Comparing performance and energy-efficiency when processing physics analysis jobs using different numbers of tasks per CPU core settings to 1 task per core setting

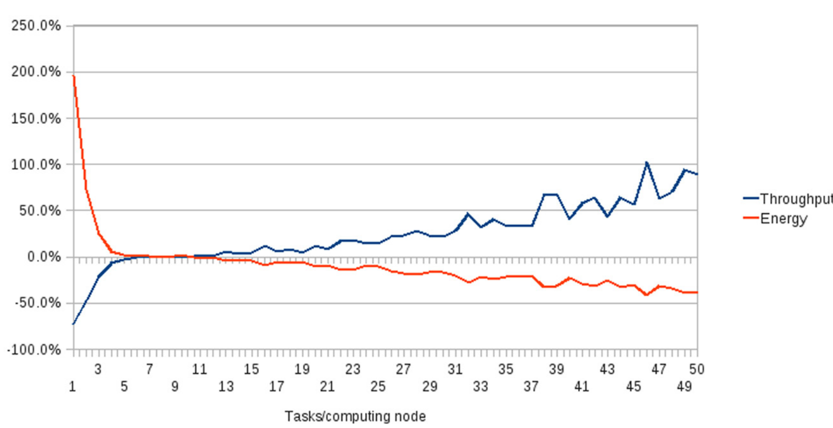
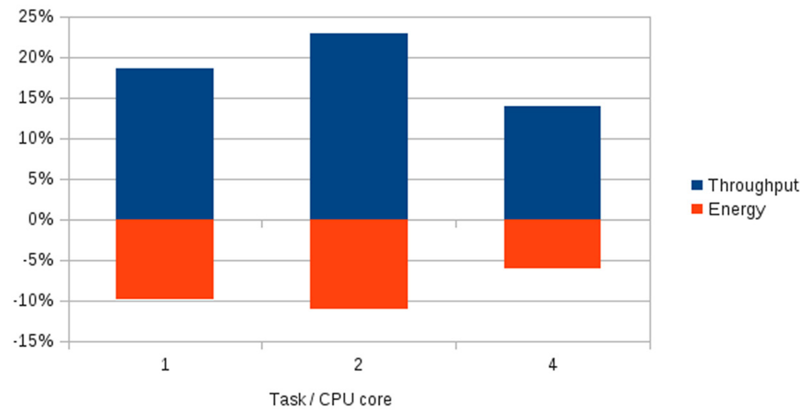


Figure 5. Comparing mixing CPU and I/O oriented physics tasks to processing first all CPU tasks



within the system becomes a bottleneck making the processor wait for some operations, causing CPU utilization to drop from the maximum 100%. Based on our tests, we can assume that I/O traffic is an obvious bottleneck.

The results are also in line with Schmenner (2010), who proposed that companies emphasizing swift even flow, i.e. they focus on speed and variability reduction, would outperform slower companies. This principle also implies that companies should focus on value adding tasks and removing of non-value adding tasks, while at the same time trying to eliminate bottlenecks in order to introduce even flow and short lead times.

Our work showed that multitasking improves efficiency in HEP data analysis but it also easily makes disk access a bottleneck. While the number of CPU cores continuously increases, I/O remains a bottleneck. Faster disks like Solid State Drives (SSD) can be used to partially remove this bottleneck, since SSD disks are clearly faster than hard disks. Our future research will also focus on this issue, and our preliminary results already confirmed this assumption (Niemi, et al., 2011). However, our results depended highly on the applications used.

Generally, value adding production systems with high throughput and short lead times have been proven to generate benefits other than pure output performance. Statistically these systems also produce better quality and less waste, and thus have a better overall environmental efficiency. Systems, which perform better operationally, also have more satisfied customers and tend to be more competitive in the market. Based on this, an optimized computing system should, in addition to reduced electricity consumption, also work more reliably and offer more computing power. When applying new optimizing methods to large scale computing resources, the results can bring remarkable savings, especially in large computing installations.

There are still several unsolved issues concerning the loading of computing resources. This chapter presented a method based on the processing of more than one task per CPU core. In this method an obvious challenge is to decide upon the number of tasks to run in parallel. Therefore, we have started to develop automated methods for dynamically finding the right settings. One method based on fuzzy control is introduced in Niemi and Hameri (2011). Another approach uses the simulation approach to test different scheduling strategies much faster than with contemporary test runs. We have developed simulator software for this purpose (Kommeri, et al., 2011).

CONCLUSION

We studied different scheduling settings with different hardware for high-energy physics computing to minimize electricity usage and maximize performance. Instead of the common practice of one task per CPU core scheduling used in many computing clusters, we tested variations of different scheduling methods based on the idea to process more tasks in parallel.

The results showed that optimizing the configuration of a workload management system by using our methods improved throughput and decreased electricity consumption. However, improvements heavily depended on the application. In a modern multicore environment running 3-4 tasks per CPU core gave the best results. The measured improvements of our methods increased throughput up to 100% and decreased energy consumption down to 40-50% compared to the standard one-task-per-core practice in scientific computing. The trade-off is that processing times of individual tasks become longer. Though in cases, such as HEP computing, in which the tasks are not time critical, the total throughput and energy-efficiency are more important. Instead, we did not find a trade-off between throughput and energy-efficiency, as in our tests the most efficient method also minimized energy consumption per processed task. As a general conclusion, our results show that information on all components of the computer, including memory usage, processor load, and I/O traffic, should be used when making scheduling decisions.

REFERENCES

Acosta, D., & Camporesi, T. (2008). *Cosmic success*. CMS Times.

Agostinelli, S. (2003). GEANT4: A simulation toolkit. *Nuclear Instruments and Methods in Physics*, *506*, 250–303. doi:10.1016/S0168-9002(03)01368-8

Antcheva, I., Ballintijn, M., & Bellenot, B. (2009). ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, *180*, 2499–2512. doi:10.1016/j.cpc.2009.08.005

Apparao, P., Iyer, R., Zhang, X., Newell, D., & Adelmeyer, T. (2008). Characterization & analysis of a server consolidation benchmark. In *Proceedings of the Fourth ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments*, (pp. 21-30). New York, NY: ACM Press.

Aziz, A., & El-Rewini, H. (2008). On the use of meta-heuristics to increase the efficiency of online grid workflow scheduling algorithms. *Cluster Computing*, *11*(4), 373–390. doi:10.1007/s10586-008-0062-y

Belady, C. (2008). Green grid data center power efficiency metrics PUE and DCiE. *The Green Grid*. Retrieved from http://www.thegreengrid.org.

Bobroff, N., Kochut, A., & Beaty, K. (2007). Dynamic placement of virtual machines for managing SLA violations. In *Proceedings of IFIP/IEEE Symposium on Integrated Network Management*. Munich, Germany: IEEE Press.

Brown, D. J., & Reams, C. (2010). Toward energy-efficient computing. *Queue*, *8*(2), 30–43.

Brucker, P. (2007). *Scheduling algorithms*. Berlin, Germany: Springer.

Bunde, D. P. (2006). Power-aware scheduling for makespan and flow. In *Proceedings of the 18th Annual ACM Symposium on Parallelism in Algorithms and Architectures*, (pp. 190–196). New York, NY: ACM Press.

Cao, J., Jarvis, S., Saini, S., & Nudd, G. (2003). Gridflow: Workflow management for grid computing. In *Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Grids*, (pp. 198 – 205). IEEE Press.

Cluster Resources. (2011). *Website.* Retrieved from http://www.clusterresources.com/pages/products/torque-resource-manager.php.

Conner, S., Link, G. M., Tobita, S., Irwin, M. J., & Raghavan, P. (2006). Energy/performance modeling for collective communication in 3-d torus cluster networks. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*. New York, NY: ACM Press.

Da Silva, P. D., Cirne, W., Brasileiro, F. V., & Grande, C. (2003). Trading cycles for information: Using replication to schedule bag-of-tasks applications on computational grids. In [IEEE.]. *Proceedings of Euro-Par*, *2003*, 169–180.

Edmonds, J. (2000). Scheduling in the dark. *Theoretical Computer Science*, *235*(1), 109–141. doi:10.1016/S0304-3975(99)00186-3

Essary, D., & Amer, A. (2008). Predictive data grouping: Defining the bounds of energy and latency reduction through predictive data grouping and replication. *Transmission and Storage*, *4*(1), 1–23. doi:10.1145/1353452.1353454

Etsion, Y., & Tsafrir, D. (2005). *A short survey of commercial cluster batch schedulers*. Technical Report 2005-13. Jerusalem, Israel: Hebrew University of Jerusalem.

Fabozzi, F., Jones, C. D., Hegner, B., & Lista, L. (2008). Physics analysis tools for the CMS experiment at LHC. *IEEE Transactions on Nuclear Science*, *55*, 3539–3543. doi:10.1109/TNS.2008.2006979

Fu, R., Ji, T., Yuan, J., & Lin, Y. (2007). Online scheduling in a parallel batch processing system to minimize makespan using restarts. *Theoretical Computer Science*, *374*(1-3), 196–202. doi:10.1016/j.tcs.2006.12.040

Ge, R., Feng, X., & Cameron, K. W. (2005). Performance-constrained distributed DVS scheduling for scientific applications on power-aware clusters. In *Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*, (p. 34). Washington, DC: IEEE Computer Society.

Goes, L. F. (2005). Anthillsched: A scheduling strategy for irregular and iterative i/o-intensive parallel jobs. In *Proceedings of Job Scheduling Strategies for Parallel Processing - JSSPP 2005*. Springer. doi:10.1007/11605300_5

Green IT Promotion Council. (2010). *Concept of new metrics for data center energy efficiency introduction of datacenter performance per energy*. Kyoto, Japan: Green IT Promotion Council.

Grosan, C., Abraham, A., & Helvik, B. (2007). Multiobjective evolutionary algorithms for scheduling jobs on computational grids. In N. Guimares & P. Isaias (Eds.), *ADIS International Conference, Applied Computing 2007*. Salamanca, Spain: ADIS.

Hameri, A.-P., & Niemi, T. (2010). Applying operations management principles on optimisation of scientific computing clusters. In *Proceedings of the 2nd Rapid Modelling Conference*. Neuchatel, Switzerland: Springer.

Hanselman, S. E., & Pegah, M. (2007). The wild wild waste: E-waste. In *Proceedings of the 35th Annual ACM SIGUCCS Conference on User Services*, (pp. 157-162). New York, NY: ACM Press.

Kappiah, N., Freeh, V. W., & Lowenthal, D. K. (2005). Just in time dynamic voltage scaling: Exploiting inter-node slack to save energy in mpi programs. In *Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*. Washington DC: IEEE Computer Society.

Karlsson, M., Karamanolis, C., & Zhu, X. (2005). Triage: Performance differentiation for storage systems using adaptive control. *Transmission and Storage*, *1*(4), 457–480. doi:10.1145/1111609.1111612

Kommeri, J., Hotakainen, R., & Niemi, T. (2011). *Modelling computation of high-energy physics jobs by using a simulator*. Paper presented at the 21st European Japanese Conference on Information Modelling and Knowledge Bases. Tallinn, Estonia.

Koole, G., & Righter, R. (2008). Resource allocation in grid computing. *Journal of Scheduling*, *11*(3), 163–173. doi:10.1007/s10951-007-0018-8

Kurowski, K., Nabrzyski, J., Oleksiak, A., & Weglarz, J. (2008). A multicriteria approach to two- level hierarchy scheduling in grids. *Journal of Scheduling*, *11*(5), 371–379. doi:10.1007/s10951-008-0058-8

Lefurgy, C., Wang, X., & Ware, M. (2007). Server-level power control. In *Proceedings of the Fourth International Conference on Autonomic Computing*. Washington, DC: IEEE Computer Society.

Li, X., Li, Z., Zhou, Y., & Adve, S. (2005). Performance directed energy management for main memory and disks. *Transmission and Storage*, *1*(3), 346–380. doi:10.1145/1084779.1084782

Liouane, N., Yahia, H., & Borne, P. (2008). Multi-objective scheduling onto heterogeneous processors system using ant system & fuzzy logic controller. *Studies in Informatics and Control*, *17*(1), 95–106.

Marwah, M., et al. (2009). Data analysis, visualization and knowledge discovery in sustainable data centers. In *Proceedings of the 2nd Bangalore Annual Compute Conference*, (pp 1-8). New York, NY: ACM Press.

Medernach, E. (2005). Workload analysis of a cluster in a grid environment. In *Proceedings of the Job Scheduling Strategies for Parallel Processing 11th International Workshop, JSSPP 2005*. Springer.

Moallem, A., & Ludwig, S. A. (2009). Using artificial life techniques for distributed grid job scheduling. In *Proceedings of the 2009 ACM Symposium on Applied Computing*, (pp. 1091–1097). New York, NY: ACM Press.

Mu'alem, A. W., & Feitelson, D. G. (2001). Utilization, predictability, workloads, and user runtime estimates in scheduling the IBM SP2 with backfilling. *IEEE Transactions on Parallel and Distributed Systems*, *12*(6), 529–543. doi:10.1109/71.932708

Mukherjee, T., Banerjee, A., Varsamopoulos, G., Gupta, S. K. S., & Rungta, S. (2009). Spatio-temporal thermal-aware job scheduling to minimize energy consumption in virtualized heterogeneous data centers. *Computer Networks*, *53*(17), 2888–2904. doi:10.1016/j.comnet.2009.06.008

Ni, L., Zhang, J., Yan, C., & Jiang, C. (2005). A heuristic algorithm for task scheduling based on mean load. In *Proceedings of the Semantics, Knowledge and Grid, International Conference*. ACM.

Niemi, T., & Hameri, A.-P. (2011). Memory-based scheduling of scientific computing clusters. *The Journal of Supercomputing*. Retrieved from http://www.springerlink.com/content/085004562g247626/.

Niemi, T., Kommeri, J., & Hameri, A.-P. (2009a). Energy-efficient scheduling of grid computing clusters. In *Proceedings of the 17th Annual International Conference on Advanced Computing and Communications (ADCOM 2009)*. Bengaluru, India: ADCOM.

Niemi, T., Kommeri, J., & Hameri, A.-P. (2011). *Using multitasking and SSD disks for optimising computing cluster energy-efficiency*. Paper presented at the First International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies, ENERGY 2011. Venice, Italy.

Niemi, T., Kommeri, J., Happonen, K., Klem, J., & Hameri, A.-P. (2009b). Improving energy-efficiency of grid computing clusters. In *Proceedings of the 4th Advances in Grid and Pervasive Computing Conference, GPC 2009*, (vol. 5529). Springer

Open, P. B. S. (2011). *Website.* Retrieved from http://www.openpbs.org.

Pinedo, M. (2008). *Scheduling: Theory, algorithms, and systems*. Berlin, Germany: Springer.

Piro, R. M., Guarise, A., Patania, G., & Werbrouck, A. (2009). Using historical accounting information to predict the resource usage of grid jobs. *Future Generation Computer Systems*, *25*(5), 499–510. doi:10.1016/j.future.2008.11.003

Platform. (2011). *Website.* Retrieved from http://www.platform.com/Products/Workload-Management.

Pradeep, P., Xiaoyun, Z., Zhikui, W., Sharad, S., & Kang, G. S. (2007). *Performance evaluation of virtualization technologies for server consolidation*. Technical report HPL-2007-59. Palo Alto, CA: Hewlett-Packard.

Prasanna, G. N. S., & Musicus, B. R. (1996). The optimal control approach to generalized multiprocessor scheduling. *Algorithmica*, *15*(1), 17–49. doi:10.1007/s004539900002

Rajan, D., & Yu, P. S. (2008). Temperature-aware scheduling: When is system- throttling good enough? In *Proceedings of the 2008 Ninth International Conference on Web-Age Information Management*, (pp. 397–404). Washington, DC: IEEE Computer Society.

Regola, N., & Ducom, J. C. (2010). *Recommendations for virtualization technologies in high performance computing*. Paper presented at CloudCom IEEE Second International Conference on Cloud Computing Technology and Science. Indianapolis, IN.

Santos-Neto, E., Cirne, W., Brasileiro, F., Lima, A., Lima, R., & Grande, C. (2004). Exploiting replication and data reuse to efficiently schedule data-intensive applications on grids. In *Proceedings of the 10th Workshop on Job Scheduling Strategies for Parallel Processing*, (pp. 210–232). ACM.

Schmenner, R. W. (2010). Looking ahead by looking back: Swift, even flow in the history of manufacturing. *Production and Operations Management*, *10*(1), 87–96. doi:10.1111/j.1937-5956.2001.tb00069.x

Shivam, P., Babu, S., & Chase, J. (2006). Active and accelerated learning of cost models for optimizing scientific applications. In *Proceedings of the 32nd International Conference on Very Large Data Bases*, (pp. 535–546). VLDB Endowment.

Sjostrand, T., Mrenna, S., & Skands, P. Z. (2006). PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 5.

Sun. (2008). *Beginner's guide to SUNTM grid engine 6.2 installation and configuration*. Santa Clara, CA: Sun Microsystems.

Tsafrir, D., Etsion, Y., & Feitelson, D. G. (2005). Modeling user runtime estimates. In *Proceedings of the Job Scheduling Strategies for Parallel Processing 11th International Workshop, JSSPP 2005*. Springer.

Tsafrir, D., Etsion, Y., & Feitelson, D. G. (2007). Backfilling using system-generated predictions rather than user runtime estimates. *IEEE Transactions on Parallel and Distributed Systems*, *18*(6), 789–803. doi:10.1109/TPDS.2007.70606

Venkatachalam, V., & Franz, M. (2005). Power reduction techniques for microprocessor systems. *ACM Computing Surveys*, *37*(3), 195–237. doi:10.1145/1108956.1108957

Viswanathan, H., Lee, E. K., Rodero, I., Pompili, D., & Parashar, M. (2011). Energy-aware application-centric VM allocation for HPC workloads. In *Proceedings of the 25th IEEE International Parallel & Distributed Processing Symposium, Workshops and Phd Forum*. Anchorage, AK: IEEE Press.

Wang, C.-M., Huang, X.-W., & Hsu, C.-C. (2009). Bi-objective optimization: An online algorithm for job assignment. In *Proceedings GPC 2009*, (pp. 223-234). Geneva, Switzerland: GPC.

Wisconsin. (2011). *Condor*. Retrieved from http://www.cs.wisc.edu/condor.

Yom-Tov, E., & Aridor, Y. (2008). A self-optimized job scheduler for heterogeneous server clusters. In *Proceedings of the 13th International Conference on Job Scheduling Strategies for Parallel Processing*, (pp. 169–187). Berlin, Germany: Springer-Verlag.

Yuan, W., & Nahrstedt, K. (2002). Integration of dynamic voltage scaling and soft real-time scheduling for open mobile systems. In *Proceedings of the 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, (pp. 105–114). New York, NY: ACM Press.

Zhang, W., Hu, J. S., Degalahal, V., Kandemir, M., Vijaykrishnan, N., & Irwin, M. J. (2004). Reducing instruction cache energy consumption using a compiler based strategy. *ACM Transactions on Architecture Code Optimization*, *1*(1), 3–33. doi:10.1145/980152.980154

Zhu, Q., Chen, Z., Tan, L., Zhou, Y., Keeton, K., & Wilkes, J. (2005). Hibernator: Helping disk arrays sleep through the winter. In Proceedings of the 20th ACM Symposium on Operating Systems Principles, (pp. 177–190). New York, NY: ACM.

ADDITIONAL READING

Barroso, L. A., & Hölzle, U. (2007). The case for energy-proportional computing. *Computer*, *40*(12), 33–37. doi:10.1109/MC.2007.443

Bohrer, P., Elnozahy, E. N., Keller, T., Kistler, M., Lefurgy, C., McDowell, C., & Rajamony, R. (2002). The case for power management in web servers. In *Power Aware Computing*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Boxwell, R. J. (1994). *Benchmarking for competitive advantage*. New York, NY: McGraw-Hill.

Carlsson, N., & Arlitt, M. (2011). Towards more effective utilization of computer systems. In *Proceeding of the Second Joint WOSP/SIPEW International Conference on Performance Engineering (ICPE 2011)*, (pp. 235-246). New York, NY: ACM Press.

e-Infrastructure Reflection Group. (2011). *e-IRG white paper 2011*. Retrieved from http://www.e-irg.eu/publications/white-papers.html.

Färe, R., Grosskopf, S., & Knox Lovell, C. A. (Eds.). (1985). *The measurement of efficiency of production*. Hingham, MA: Kluwer.

Gandhi, A., Harchol-Balter, M., Das, R., & Lefurgy, C. (2009). Optimal power allocation in server farms. In *Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 2009)*, (pp. 157-168). New York, NY: ACM Press.

Goldratt, E. M., & Cox, J. (1984). *The goal. Croton-on-Hudson*. NY: North River Press.

Graefe, G. (2008). Database servers tailored to improve energy efficiency. In *Proceedings of the 2008 EDBT Workshop on Software Engineering for Tailor-Made Data Management, SETMDM 2008*, (pp. 2428). New York, NY: ACM Press.

Hameri, A.-P., & Nihtilä, J. (1998). Computerized product process: Measurement and continuous improvement. *Research in Engineering Design*, *10*(3), 166–177. doi:10.1007/BF01607158

Härder, T., Hudlet, V., Ou, Y., & Schall, D. (2011). Energy efficiency is not enough, energy proportionality is needed! In *Proceedings of the 16th International Conference on Database Systems for Advanced Applications, DASFAA 2011*, (pp. 226239). Springer-Verlag.

Harizopoulos, S., Liang, V., Abadi, D. J., & Madden, S. (2006). Performance tradeoffs in read-optimized databases. In *Proceedings of the 32nd International Conference on Very Large Data Bases, VLDB 2006*, (pp. 487498). VLDB Endowment.

Harizopoulos, S., Shah, M. A., Meza, J., & Ranganathan, P. (2009). *Energy efficiency: The new holy grail of data management systems research*. Paper presented at the 4th Biennial Conference on Innovative Data Systems Research (CIDR). Asilomar, CA.

Herrero, I., & Salmeron, J. L. (2005). Using the DEA methodology to rank software technical efficiency. *Communications of the ACM*, *48*(1), 101–105. doi:10.1145/1039539.1039546

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly*, *28*(1), 75–105.

Hölzle, U., & Weihl, B. (2006). *PSU white paper*. Retrieved from http://www.psu.edu.

Hopp, W. J., & Spearman, M. L. (1996). *Factory physics*. Chicago, IL: Irwin.

Lang, W., Kandhan, R., & Patel, J. M. (2011). Rethinking query processing for energy efficiency: Slowing down to win the race. *A Quarterly Bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering*, *34*(1), 1223.

Lang, W., & Patel, J. (2009). *Towards eco-friendly database management systems*. Paper presented at the 4th Biennial Conference on Innovative Data Systems Research (CIDR). Asilomar, CA.

Lang, W., Patel, J. M., & Shankar, S. (2010). Wimpy node clusters: What about non-wimpy workloads? In *Proceedings of the Sixth International Workshop on Data Management on New Hardware, DaMoN 2010*, (pp. 4755). New York, NY: ACM Press.

Li, Z., Grosu, R., Sehgal, P., Smolka, S., Stoller, S., & Zadok, E. (2011). On the energy consumption and performance of systems software. In *Proceedings of the 4th Annual International Conference on Systems and Storage (SYSTOR 2011)*. New York, NY: ACM Press.

Meisner, D., Gold, B. T., & Wenisch, T. F. (2009). PowerNap: Eliminating server idle power. In *Proceeding of the 14th international Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS 2009*, (pp. 205-216). Washington, DC: ACM Press.

Meza, J., Shah, M. A., Ranganathan, P., Fitzner, M., & Veazey, J. (2009). Tracking the power in an enterprise decision support system. In *Proceedings of the 14th ACM/IEEE International Symposium on Low Power Electronics and Design, ISLPED 2009*, (pp. 261266). New York, NY: ACM Press.

Patel, C., & Shah, A. (2005). *Cost model for planning, development and operation of a data center*. HP Laboratories Technical Report, HPL-2005-107R1. Palo Alto, CA: HP.

Rasmussen, N. (2007a). *Calculating total cooling requirements for data centers*. White Paper #25. Washington, DC: American Power Conversion.

Rasmussen, N. (2007b). *Avoiding costs from oversizing data center and network room infrastructure*. White Paper #37. Washington, DC: American Power Conversion.

Tsirogiannis, D., Harizopoulos, S., & Shah, M. (2010). Analyzing the energy efficiency of a database server. In *Proceedings of the 2010 International Conference on Management of Data, SIGMOD 2010*. SIGMOD.

Wilf, H. S. (1986). *Algorithms and complexity*. Englewood Cliffs, NJ: Prentice-Hall.

Xu, Z., Tu, Y.-C., & Wang, X. (2010). Exploring power performance tradeoffs in database systems. In *Proceedings of the 26th International Conference on Data Engineering (ICDE)*, (pp. 485-496). IEEE Press.

KEY TERMS AND DEFINITIONS

Computing Jobs and Tasks: A task is the smallest elementary entity when work is processed. The task starts, retrieves/reads its possible input file, processes the data, and possibly writes its output file. A job is a collection of tasks.

Energy Efficiency: Energy efficiency means how efficiently energy is used. For example how many of similar tasks or jobs that can be processed by using a fixed amount of electricity.

High-Energy Physics Computing: High Energy Physics is a discipline, which studies sub-atomic particles such as electrons, protons, and neutrons. The name "high-energy" comes from the fact that one needs very high energy in particle accelerators to separate these elementary particles apart. A lot of computing power is needed to analyze data collected by detectors in particle accelerators.

Load: Measurement of utilization of a computer. It indicates how many processes are waiting for CPU becoming available.

Scheduling: Scheduling means how to allocate computing tasks to computing nodes.

Scientific Computing Cluster: A group of computers that appears as one. They are connected via high a speed local area network. A cluster consists of at least one frontend and one too many computing nodes. The frontend is a machine that acts as a gateway and moderator for the cluster. Computing nodes are computers that do the actual computing.

Throughput: Throughput means the number of similar tasks or jobs that can be processed in a time unit.

Workload Management System: A system that keeps track of activity in a computing cluster and schedules tasks into computing nodes based on this knowledge.

ENDNOTES

1. See Cluster Resources (2011).
2. See Open PBS (2011).
3. See Platform (2011).
4. See Wisconsin (2011).

This work was previously published in Energy-Aware Systems and Networking for Sustainable Initiatives, edited by Naima Kaabouch and Wen-Chen Hu, pp. 1-19, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 104
Organic Solar Cells Modeling and Simulation

Mihai Razvan Mitroi
Polytechnic University of Bucharest, Romania

Laurentiu Fara
Polytechnic University of Bucharest, Romania & Academy of Romanian Scientists, Romania

Andrei Galbeaza Moraru
Polytechnic University of Bucharest, Romania

ABSTRACT

Modelling and simulation of organic (polymer, dye sensitized, and nanotube) solar cells is discussed. High J-V theoretical curves, the calculation of key parameters, and also the relative influence of different parameters on the cell operation, are evidenced and analyzed. On this basis, the authors obtain information on the optimization of the solar cell design and manufacturing.

MODELING AND SIMULATION OF ORGANIC CELLS BASED ON POLYMERS

Introduction

The polymer-based photovoltaic devices present some important advantages, such as the low cost and the easy manufacturing from thin films by chemical/physical vapour deposition, screen-printing, or casting. The band gap of the films can be adjusted by chemical synthesis to convenient values and the carrier mobility can reach 10 $cm^2/V \cdot s$ (Dimitrakopoulos & Mascaro, 2001; Kwok, 2003). Therefore, the polymer-based photovoltaic cells are competitive with those based on amorphous silicon, so that the interest for both the properties of the polymer materials and the characteristics of the polymer-based cells increases.

The first structure was a Schottky diode type structure, which has a conversion efficiency under 1%. In 1986, Tang (1986) introduced the planar donor-acceptor heterojunction, which has a higher conversion efficiency. Hiramoto, Fujiwara, and Yokoyama (1991) developed this idea. Tang introduced the concept of Bulk Heterojunction—BH. Xue et al. (2005) introduced the concept of Hybrid Heterojunction (HH) with higher efficiency 5%.

DOI: 10.4018/978-1-4666-1945-6.ch104

Solar cells based on organic polymers used photoinduced transfer of electrons from semiconductor polymers, donor / acceptor polymers or acceptor molecules (such as C_{60}). Pairs of layers polymer / fullerene have a poor conversion efficiency of solar cell. Nanomorfological control at nanoscale of the separate regions in an interlaced network (BH) significantly increased the conversion efficiency of solar energy cells made of MDMO-PPV/C_{60}. Typical dimensions of the separate regions must be smaller than the exciton diffusion length. On the other hand, bicontinue percolation paths for the transport of charge carriers at the electrodes must be provided in order to increase the transport of charge carriers in organic and polymeric materials, and for mesoscopic order to improve crystallinity. So, an interlaced network at nanoscale with crystalline order of the both constituents is formed (these ones belong to individual sub-networks). Such a network is a convenient structure for the active layer of polymer PV devices. At the same time, forbidden bands of the photoactive layer materials should be chosen in order to absorb more light in solar radiation spectrum.

Using a mixture of MDMO-PPV soluble fullerene derivative (for example PCBM), it is possible to obtain solar cells with a 2.5% conversion efficiency.

It was shown (Shaheen, et al., 2001) that a conversion efficiency of 2.5% (for AM 1.5) can be obtained using chlorobenzenes as a solvent for the deposition by centrifuging in the mass ratio of 1:4 for the MDMO- PPV: PCBM. Using chlorobenzenes instead of toluene nanomorfology changes were remarked and the efficiency conversion increases 3 times. Such BH solar cells contain 80% PCBM. However, the MDMO-PPV polymer could be the main light absorber from these solar cells, because PCBM does not achieve almost no absorption in the visible and near infrared region. So an increase in volume concentration of MDMO-PPV was necessary for a better absorption of sunlight.

The electron mobility from pure PCBM is proved to be higher $(\sim 10^{-3}\ \mathrm{cm}^2/\mathrm{Vs})$ compared to the holes mobility from the MDMO-PPV $(\sim 10^{-4}\ \mathrm{cm}^2/\mathrm{Vs})$ and the holes mobility in the mixture increases with the increasing of fullerene mass, despite of the fact that the addition of fullerene introduces more defects (which would reduce the mobility).

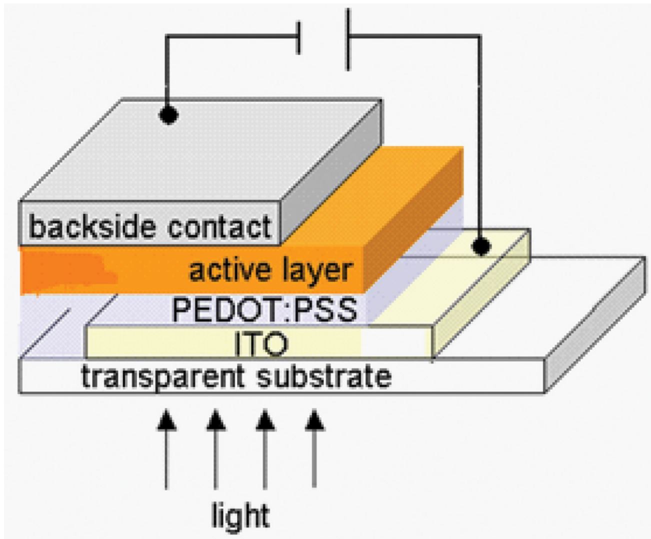
By replacing in PCBM the C_{60} fullerene with C_{70} one, the HOMO-LUMO transition would be easier and would increase the absorption of light. An improved efficiency of BH solar cells in isomeric mixture of MDMO-PPV and C_{70} derivatives was obtained (Wienk, et al., 2003). Mixtures of this kind are suitable when they are prepared from a solution of o-dichlorobenzene (ODCB) and solar cells have a conversion efficiency of $\eta = 3\ \%$ for PM 1.5.

Operating Principle of Polymer-Based Solar Cell

The process of conversion of light into electricity using a organic solar cell (Figure 1) can be plotted by the following steps: 1) absorption of a photon determining the formation of an excited state, that is, the creation of an exciton (electron-hole pair); 2) exciton diffusion to a region where dissociation occurs; 3) charge transport within the organic semiconductor to the respective electrodes.

The obtained photo-voltage is a direct result of the departure from thermodynamic equilibrium state after the movement of charge carriers, generated by solar radiation under the action of local fields, which in the case of organic semiconductors is due to the change in chemical composition. Because of the large band gap in organic materials, only a small portion of the incident solar light is absorbed. A band gap of 1.1 eV (1100 nm) is able of absorbing 77% of the solar irradiation on Earth (Nunzi, 2002). But, the majority of semiconducting polymers have band gaps higher than 2 eV (620 nm), which limits the possible harvest-

Figure 1. Schematic representation of standard bulk heterojunction solar cell



ing of solar photons to about 30% (Nunzi, 2002). On the other hand, because the absorption coefficients of organic materials are higher than 10^5 cm^{-1}, only 100 nm thickness is enough to absorb most of the photons when a reflective back contact is used.

The primary photoexcitations in organic materials do not influence directly the release of charge carriers, but determine to obtain the electron-hole pairs (it is estimated that only 10% of photoexcitations leads to the free charge carriers in conjugated polymers). For an efficient dissociation of excitons, strong electric fields are necessary.

At an interface, where sudden changes of the potential energy occur, strong local electrical fields are possible ($E = -grad\ U$). Photoinduced charge transfer can occur when the exciton has reached such an interface within its lifetime. Therefore, the exciton diffusion length limits the thicknesses of the double layers. Exciton diffusion length should be the same order of magnitude with the separation distance between donor and acceptor fields. Otherwise, the excitons recombine radiatively or non-radiatively before reaching the interface and their energy no longer contributes to solar energy conversion. Exciton diffusion length in organic polymers is usually around 10÷20 nm.

In the materials obtained by combining conjugated polymers with the acceptor materials, as fullerenes, the efficiency of converting the photogenerated excitons to free charge carriers increases. By the absorption of light in a donor an excited state from which the electron can be transferred to the LUMO unoccupied orbital of the acceptor arises. Following the transfer, the hole that appears remains in the donor material. Measurements showed that the photoinduced charge transfer in such a compound occurs in $45\ fs$, so in a much smaller time than other similar processes (e.g. photoluminescence transition lasts around 1 ns) (Brabec, et al., 2001).

Charge carriers are extracted from the device by two selective contacts. A contact of Indium and Tin Oxide (ITO) with a Fermi level of about 4.8eV matches with the HOMO levels of most conjugated polymers (contact holes), and a contact of aluminum with a Fermi level of roughly 4.3.eV matches with the LUMO of the PCBM acceptor (contact for electrons).

Theoretical Models

A number of models were developed aimed to the modeling and simulation of the organic solar cells based on polymers. Many of these models based on the diode-like equivalent circuit of a solar cell are analysing the $J-V$ characteristics; they can simulate the cell behaviour and/or evaluate their macroscopic parameters (Aernouts, 2006; Marlein & Burgelman, 2007; Marlein, Decock, & Burgelman, 2009; Greulich, et al., 2010; Chuan, et al., 2011).

In the absence of illumination typical current density-voltage characteristics of a solar cell is shown in Figure 2a; in the presence of illumination it is shown in Figure 2b.

In the following a modelling and simulation analyses is presented (Mitroi, Iancu, Fara, & Ciurea, 2010).

To represent a solar cell under illumination, an equivalent electronic circuit can be drawn as shown in Figure 3. It consists basically from a constant current source parallel to a diode. This current source is supposed to be bias independent and represents the current J_L, generated in the device due to the incoming light. Its value might thus vary according to the intensity of the illumination. A series and parallel resistance, R_s and R_p respectively, are additionally included to account for non-ideal behavior.

Under illumination the total current density is:

$$J = J_s \left\{ \exp\left[\frac{e\left(V - R_s J\right)}{\gamma k_B T} \right] - 1 \right\} + \frac{V - R_s J}{R_p} - eG\mu\tau \left(\frac{V_{bi} - V}{L} \right) \cdot \left[1 - \exp\left(\frac{L^2}{\mu\tau\left(V_{bi} - V\right)} \right) \right] \quad (1)$$

where J_s is the saturation current density, $e > 0$ is the elementary charge, V is the applied voltage, R_s and R_p are the linear series and derivation resistances, γ is the ideality factor, k_B is the Boltzmann constant, and T is the cell temperature, G is the pair generation rate, V_{bi} is the bias voltage, V is the voltage on a external resistance R_{ext}, L is the thickness of the active layer, and the product $\mu\tau$ is given by the relation:

$$\mu\tau = \mu_p\tau_p + \mu_n\tau_n, \quad (2)$$

where μ_p, μ_n, and τ_p, τ_n are the hole and electron mobilities and lifetimes, respectively. This product obviously depends on the recombination and carrier transport processes.

Figure 2. Current density-voltage characteristic in the absence of illumination

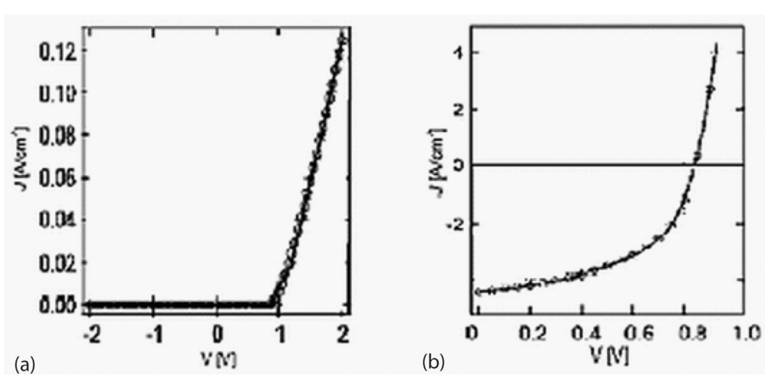
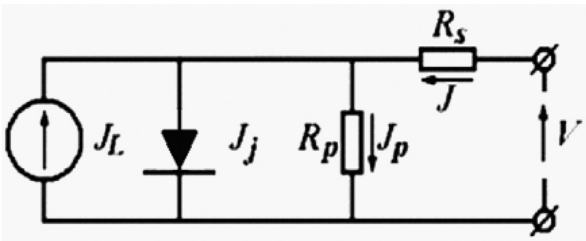


Figure 3. The diode-like equivalent circuit of a solar cell under illumination



The effect of R_s and R_p on the illuminated *J-V* characteristics is shown in Figure 4.

This resistances act to reduce the fill factor, whereas very high values of R_s and very low values of R_p can also reduce J_{sc} and V_{oc}, respectively.

From the Formula (1) the following relationships results:

$$J_{sc} = J_s \left\{ \exp\left[\frac{e\left(-R_s J_{sc}\right)}{\gamma k_B T} \right] - 1 \right\} - \frac{R_s J_{sc}}{R_p} - eG\mu\tau \left(\frac{V_{bi}}{L} \right) \cdot \left[1 - \exp\left(\frac{L^2}{\mu\tau V_{bi}} \right) \right] \tag{3}$$

and

$$J_s \left\{ \exp\left[\frac{eV_{oc}}{\gamma k_B T} \right] - 1 \right\} + \frac{V_{oc}}{R_p} - eG\mu\tau \left(\frac{V_{bi} - V_{oc}}{L} \right) \cdot \left[1 - \exp\left(\frac{L^2}{\mu\tau \left(V_{bi} - V_{oc} \right)} \right) \right] = 0 \tag{4}$$

Equations (3) and (4) allow the calculation of the short-circuit current density J_{sc} and the open circuit voltage V_{oc}, respectively. Based on *J-V* theoretical curve, the Fill Factor (*FF*), and the energy conversion efficiency (η) are defined as:

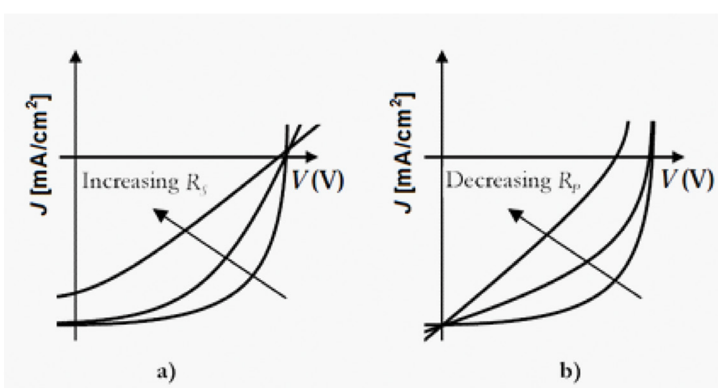
1. The Fill Factor (FF):

$$FF = \frac{V_M J_M}{V_{oc} J_{sc}} = \frac{P_M}{V_{oc} J_{sc}}, \tag{5}$$

where $P_M = V_M J_M$ is the maximum power density. V_M and J_M are the corresponding voltage and current density and they are determined from the $J-V$ curve.

2. The power conversion efficiency (□)

Figure 4. Parallel and series resistance influence on the J-V characteristic of the cell



$$\eta = \frac{P_M}{P_L} \cdot 100\ \% = \frac{V_M J_M}{P_L} \cdot 100\ \%, \qquad (6)$$

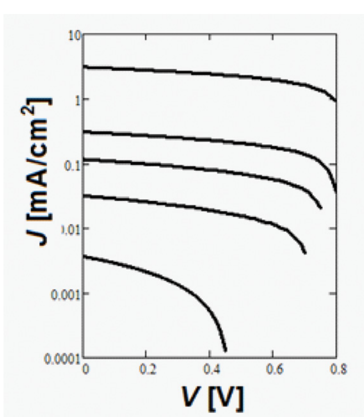
where P_L is the incident light flux power density.

Results and Discussion

The analyzed model allows the determination of the main parameters of the cell: the saturation current density J_s, the series and derivation (parallel) linear resistances R_s and R_p, the ideality factor γ, the product $\mu\tau$, the pair generation rate G, the built-in voltage V_{bi}, the short-circuit current density J_{sc}, the open-circuit voltage V_{oc}, the fill factor FF and the power conversion efficiency η. These parameters describe both the physical nature of the active layer and the technological characteristics of the cell.

From the Equation (3), one can see that J depends on seven parameters: J_s, R_s, R_p, μ, τ, γ, G, and V_{bi}. These parameters can be obtained by fitting the experimental data for the $J - V$ characteristics with this formula. These values were then introduced in the equation (8.1), and theoretical curves can be drawn (see Figure 5). Theoretical curves are obtained using the values of J_M and V_M. Then by solving the system of Equations (3) and (4), Jsc and Voc could be obtained, and by using the Equations (5) and (6) FF and η could be calculated.

Figure 5. J – V characteristics for the investigated cell



The dependences of the fill factor and the power conversion efficiency on the thickness of the active layer highlights the following results (Mitroi, Iancu, Fara, & Ciurea, 2010): the fill factor has a maximum at $L \approx 120\,\mathrm{nm}$, namely $FF_{max} \approx 44.5\ \%$, while the power conversion efficiency has a maximum of $\cdot_{max} \approx 1.22\ \%$ at $L \approx 230$ nm.

Conclusion

This model highlights the properties to be considered for improvement of photovoltaic cells based on organic polymers. Results obtained using this model are consistent with those reported in the literature (Nelson, 2003; Dennler, et al., 2006; Greulich, et al., 2010).

The model allows the determination of all the main parameters of the cell: the saturation current density J_s, the series and derivation (parallel) linear resistances R_s and R_p, the ideality factor γ, the product $\mu\tau$, the pair generation rate G, the bias voltage V_{bi}, the short-circuit current density J_{sc}, the open circuit voltage V_{oc}, the fill factor FF, and the conversion efficiency η. These parameters describe both the physical nature of the active layer and the technological characteristics of the cell. In addition, this model can determine the influence of various parameters on the total current density, fill factor and, importantly, the conversion efficiency.

The appearance of a maximum of the conversion efficiency leads to the optimization of such solar cells design and manufacturing.

MODELLING AND NUMERICAL SIMULATION OF DYE-SENSITIZED SOLAR CELLS

Introduction

Dye-Sensitized Solar Cells (DSSC) are based on sensitized nanostructures or mesoporous metal oxides. DSSC have attracted considerable attention starting from the work of O'Regan and Grätzel (1991; Grätzel, 2003).

Dye Sensitized Solar Cell (DSSC) is the only one that can offer both the flexibility and transparency.

Although, good results were obtained using DSSC both with nanostructures of TiO_2 and ZnO electrodes, ionic liquid electrolytes, carbon nanotubes, graphene and solid state DSSC (Cheng & Hsieh 2010**;** Nazeeruddin, et al., 2005; Gao, et al., 2008; Giannouli, et al., 2010; Meng, et al., 2003; Zhang & et al., 2004), maximum efficiency of such devices was only 11% (Chiba, et al., 2006), in comparison with other reported results (Zhao, et al., 1998; Contreras, et al., 1999; King, et al., 2007). These cells have potentially reduced fabrication costs in comparison with conventional silicon-based solar cells, due to the low costs of materials and the facility of the fabrication process. An optimization of DSSC requires a better insight into the interrelated processes of transport and accumulation of electrons in the mesoporous oxide phase and recombination of electrons with electron acceptors (Ferber & Luther, 2001; Knodler, et al., 1993; Smestad, Bignozzi, & Argazzi, 1994). Charge generation, transport, and recombination in DSSC are generally described by a continuity equation, in which transport is assumed to be purely diffusive, i.e. driven by concentration gradients only. However, time-dependent characterization techniques have later shown that charge transport through the TiO_2 film is impaired by multiple trapping/detrapping events (Bisquert & Vikhrenko, 2004).

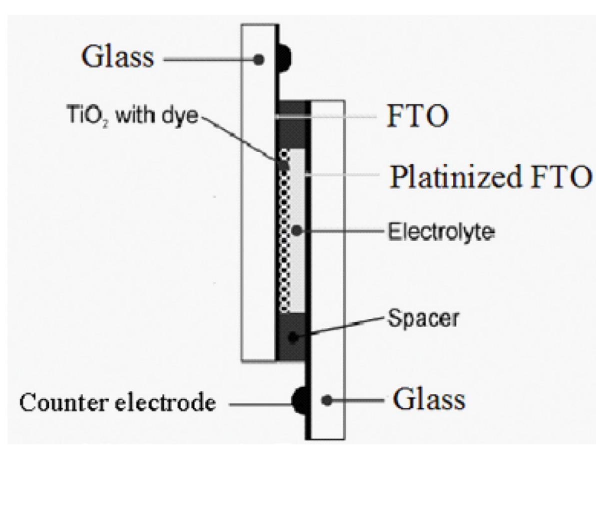
Theoretical Model

The device structure of the DSSC used in our model is presented in Figure 6. Sensitive dye solar cell is formed by a TiO_2 electrode nanopores; the pores are filled with an iodine redox electrolyte / triodura, coated with a dye that provides absorbing and transferring light task.

The two glass electrodes are made of transparent conducting electrode (fluorine-doped tin oxide, FTO). The front electrode is a transparent conductive oxide glass (FTO) coated with nanoporous TiO_2 covered with a monolayer of the Ruthenium-complex dye; the counter electrode is a FTO glass coated with a thin layer of platinum (~5 nm), where it regenerates the oxidized mediator. The gap between the two electrodes is filled by an electrolyte containing an iodide/tri-iodide (I^- / I_3^-) redox couple. Besides two different anions, the electrolyte also consists of cations being introduced by ionic liquid, usually by propylmethyl-imidazolium iodide. The cations provide electro-neutrality of the electrolyte; nevertheless they are not involved in the charge transport process in dye-sensitized solar cells.

Figure 7 shows the mechanism of a traditional wet-type DSSC containing redox couples in elec-

Figure 6. Device structure of the dye-sensitized solar cell



trolyte. Under illumination, the dye molecules become excited and initial charge separation occurs in a femtosecond time regime by injection of an electron from the dye into the conduction band of the TiO_2. This electron is transported to the external load via the nanostructured TiO_2 and TCO, while the positively charged photo-excited (S^+) dye is restored to its ground state (S) in a nanosecond time regime by receiving an electron from the iodide (I^-) (Wei, D., 2010).

Regeneration of iodide ions, which are oxidized in this reaction to tri-iodide, is achieved at a platinised counter electrode, according to Equations (7), (8), (9), (10), (11), and (12):

$$TiO_2\left|S + h\nu \rightarrow TiO_2\right|S^* \quad (7)$$

$$TiO_2\left|S^* \rightarrow TiO_2\right|S^+ + e \quad (8)$$

$$TiO_2\left|S^+ + e \rightarrow TiO_2\right|S \quad (9)$$

$$TiO_2\left|S^+ + \frac{3}{2}I^- \rightarrow TiO_2\right|S + \frac{1}{2}I_3^- \quad (10)$$

$$\frac{1}{2}I_3^- + e \rightarrow \frac{3}{2}I^- \quad (11)$$

$$I_3^- + 2e \rightarrow 3I^- \quad (12)$$

In these equations, *S* represents the dye sensitizer.

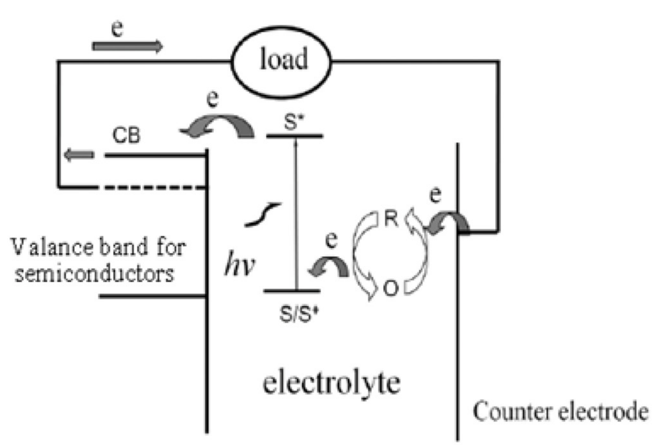
Iodide is subsequently regenerated by the reduction of tri-iodide (I_3–) at the counter electrode (Equation 12) to complete the circuit.

The Pt on the counter electrode acts as a catalyst for the reduction of tri-iodide. So far, the highest reported efficiency for a small area (< 0.2 cm^2) DSSC is 11.3%, using acetonitrile electrolytes (Grätzel, 2006) while great effort is still devoted to enhance the recorded efficiency.

Electrical Model

We will present a electric model of dye-sensitized solar cell where the active layer is modelled as a one dimensional pseudo-homogeneous layer. The electro active particles included in the model are generated by free electrons injected in nanoporous TiO_2; iodide and tri-iodide ions and cations are introduced into the cell together with electrolyte. The electrical model is based on continuity, transport, and Poisson's non-linear differential

Figure 7. The energy level scheme and the operating principle of a DSSC



equations and permits the steady-state calculation of the complete *J-V* characteristics, concentration of electro active species and their current densities in dye-sensitized solar cells (Oda, Tanaka, & Hayase, 2006; Ferber, Stangl, & Luther, 1998; Soedergren, Hagfeldt, Olsson, & Lindquist, 1994; Smestad, Bignozzi, & Argazzi, 1994).

The first important equation that connects generations and recombinations of dye-sensitized solar cell is a continuity equation for conduction band electrons:

$$-\frac{1}{e}\frac{dJ}{dx} = G(x) - R(x) \tag{13}$$

Here, *e* is the elementary charge, *x* represents the position inside the TiO_2 film thickness in interval [0, *d*], *G(x)* is the electron injection rate, determined from the optical model, *R(x)* is the recombination rate.

We assume that only electrons from the conduction band can recombine with tri-iodide in the electrolyte and that the recombination rate is linear in $n(x) - \bar{n}$:

$$R(x) = \frac{n(x) - \bar{n}}{\tau} \tag{14}$$

Here, τ is the lifetime of the conduction band electrons and $\bar{n}$ is the electron number density at equilibrium in the dark.

The movement of all four charged species in pseudo-homogeneous medium could be described by transport equations:

$$-\frac{1}{e}\frac{dJ_e(x)}{dx} = D_e\frac{d\,n_e(x)}{dx} + \mu_e\,n_e(x)E(x) \tag{15}$$

$$-\frac{1}{e}\frac{dJ_{I^-}(x)}{dx} = D_{I^-}\frac{d\,n_{I^-}(x)}{dx} + \mu_{I^-}\,n_{I^-}(x)E(x) \tag{16}$$

$$-\frac{1}{e}\frac{dJ_{I_3^-}(x)}{dx} = D_{I_3^-}\frac{d\,n_{I_3^-}(x)}{dx} + \mu_{I_3^-}\,n_{I_3^-}(x)E(x) \tag{17}$$

$$-\frac{1}{e}\frac{dJ_c}{dx} = -D_c\frac{d\,n_e(x)}{dx} + \mu_c\,n_c(x)E(x) \tag{18}$$

where D_e, D_{I^-}, $D_{I_3^-}$, and D_c are the diffusion coefficients, n_e, n_{I^-}, $n_{I_3^-}$, and n_c are electron, iodide, tri-iodide and cations concentrations, μ_e, μ_{I^-}, $\mu_{I_3^-}$, and μ_c are the mobilities of the individual charged spices, while $E(x)$ is the electrical field. Relation between diffusion coefficient and mobility of the same charged species could be described with the Einstein relation:

$$D = \frac{k_B T}{e}\mu \tag{19}$$

where k_B is the Boltzman constant and *T* is thermodynamic temperature.

Optical Model

The optical model is based on the Lambert-Beer law to determine the electron injection rate $G_e(x)$. The rate of incident photons flux is given by:

$$G_\lambda(x) = \int_{\lambda_1}^{\lambda_2} \alpha(\lambda)\,\varphi(\lambda)\exp\left(-\alpha(\lambda)\cdot x\right)d\lambda \tag{20}$$

Where λ is the wavelength of radiation, limits λ_1, λ_2 are imposed by the characteristics of absorptance α (λ) of dye used, ϕ (λ) is the incident photon flux density, and *x* is the distance from the illuminated surface of the cell to the point found inside the cell.

The incident photon flux density is reduced due to absorptance, reflectance and glass substrates

(Smestad, Bignozzi, & Argazzi, 1994), and relation (20) becomes:

$$G_{\lambda}(x) = \beta \cdot G_{\lambda}(x) \tag{21}$$

Here β is a factor that shapes the reflectance and absorptance.

The electron injection rate $G_e(x)$ is obtained multiplying the relation (21) with the injection efficiency η_{inj} :

$$G_e(x) = \eta_{inj} \cdot \beta \cdot G_{\lambda}(x) \tag{22}$$

If the quasi-equilibrium approximation is used (one single quasi-Fermi level for conduction band and trapped electrons), then, under steady-state conditions, the continuity equation does not include terms due to trapping/detrapping.

Boundary Conditions

Boundary conditions, in the integral form, are necessary to solve a coupled set of non-linear differential equations.

Conditions of preserving the total number of particles for each type of charge are:

$$\int_0^d n_c(x)dx = n_c^0 \cdot d \tag{23}$$

$$\int_0^d \left(\frac{1}{3} n_{I^-}(x) + n_{I_3^-}(x) \right) dx = \left(\frac{1}{3} n_{I^-}^0 + n_{I_3^-}^0 \right) \cdot d \tag{24}$$

$$\int_0^d \left(\frac{1}{3} n_{I^-}(x) + \frac{1}{2} n_e(x) \right) dx = \left(\frac{1}{3} n_{I^-}^0 + \frac{1}{2} n_e^0 \right) \cdot d \tag{25}$$

The boundary conditions which assure electroneutrality of solar cells are:

$$\int_0^d \left(n_e(x) + n_{I^-}(x) + n_{I_3^-}(x) \right) dx = \left(n_e^0 + n_{I^-}^0 + n_{I_3^-}^0 \right) \cdot d \tag{26}$$

$$\int_0^d \left(n_e(x) + n_{I^-}(x) + n_{I_3^-}(x) - n_c(x) \right) dx = 0 \tag{27}$$

The electric field at both boundaries equals zero, i.e.:

$$E(0) = E(d) = 0 \tag{28}$$

The electron current density at the front TCO $(x = 0)$ equals to external current density:

$$J_e(0) = J_{ext} \tag{29}$$

No electron transfer exists at $x = d$ since there is no TiO_2 layer and only a tiny bulk electrolyte layer exists:

$$J_e(d) = 0 \tag{30}$$

The opposite is true for iodide and tri-iodide ions where the iodide and tri-iodide current densities at $x = 0$ equals 0.

The cell's voltage, V, is calculated by the following relation:

$$V = \frac{1}{e}\left(E_F^n(0) - E_R(d) \right) \tag{31}$$

Here, $E_F^n(0)$ is the Fermi level of TiO_2 at the TCO-TiO_2 interface which is determined from the concentration of free electrons at TCO-TiO_2 interface $n_e(0)$:

$$E_F^n(0) = E_{CB} + k_B \cdot T \cdot \ln\left(\frac{n_e(0)}{N_{CB}}\right) \quad (32)$$

where N_{CB} is the effective density:

$$N_{CB} = 2 \cdot \left(\frac{2 \cdot \pi \cdot m_e^* \cdot k_B \cdot T}{h^2}\right)^{\frac{3}{2}} \quad (33)$$

and m_e^* is the effective electron mass, h is Planck's constant, and E_R (d) is the redox potential at counter electrode (the deviation of E_R from equilibrium value due to diffusion limitation is not taken into account).

Results and Discussion

For the numerical simulation, incident photon flux density was calculated using the global distribution of the solar spectrum type AM1.5, between the limits $\lambda_1 = 300\,nm$ and $\lambda_2 = 800\,nm$, because ruthenium was considered as dye. It was chosen because it is one of the dyes commonly used, and has values of $\alpha(\lambda)$ during the (λ_1, λ_1) range (Ferber, Stangl & Luther, 1998). The model permits the steady-state calculation of the complete *J-V* characteristic, electro active spices densities and their current densities in dye-sensitized solar cell. Such a current-voltage characteristic for illuminated DSSC is shown in Figure 8.

Using the simulated *J-V* characteristics could be determined the following parameters:

1. The fill factor:

$$FF = \frac{V_M\, J_M}{V_{oc}\, J_{sc}} = \frac{P_M}{V_{oc}\, J_{sc}}, \quad (34)$$

where $P_M = V_M\, J_M$ is the maximum power density; V_M and J_M are the corresponding voltage and current density and they are determined from the *J-V* curve.

2. The conversion efficiency:

$$\eta = \frac{P_M}{P_L} \cdot 100\ \% = \frac{V_M J_M}{P_L} \cdot 100\ \%, \quad (35)$$

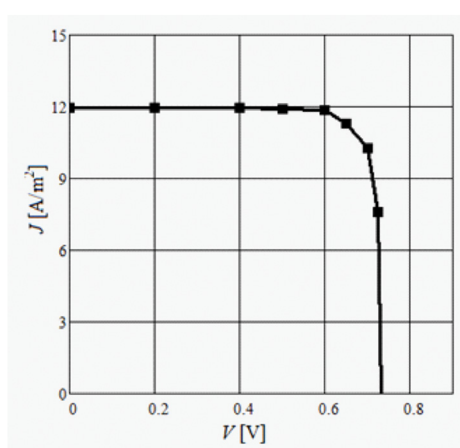
where P_L is the incident light flux power density.

Conclusion

The method allow simulated *J-V* characteristics of dye-sensitized solar cells and determination of all main parameters characterizing solar cells: the short-circuit current density J_{sc}, the open circuit voltage V_{oc}, the fill factor *FF* and the conversion efficiency η. These parameters describe both the physical nature of the active layer and the technological characteristics of the cell.

- The used method also enable to determine the influence of the different parameters of the cells on the total current density, the fill factor and, the most important, the conversion efficiency.

Figure 8. The typical simulated J-V characteristics for illuminated DSSC



- Internal electric field has low values, up to 1 mV / cm, confirming that transport tasks are carried out mainly by diffusion.
- The results based on such a model (Wenger, et al., 2009; Berginc, et al., 2009; Wenger, 2010) are consistent with theoretical results reported by different authors (Oda, Tanaka, & Hayase, 2006; Ferber, Stangl, & Luther, 1998; Oda, Tanaka, & Hayase, 2006; Ferber, Stangl, & Luther, 1998; Green, et al., 2011; Bowers, et al., 2010; Lai, et al., 2010; Agnaldo, et al., 2009; Hardin, et al., 2009; Deb, 2004; Yong, et al., 2008; Tian, et al., 2011).

NUMERICAL SIMULATION OF HIGH-ORDER TIO_2 NANOTUBE DYE-SENSITIZED SOLAR CELLS

Introduction

The electron-collecting layer in a DSSC is typically a 10 mm thick nanoparticle film, with a three-dimensional network of interconnected 15÷20 nm sized nanoparticles (Gratzel, 2004). The slow percolation of electrons (through the three-dimensional network) and low absorption of photons are two major factors which are limiting the conversion efficiency of DSSC cells using nanoparticles (Law, et al., 2005).

By using the nanotubes we could obtain a considerable increase in the photo- conversion of the DSSC. This is due to the improvement of the electrons transport in comparison to the nanoparticles one (Adachi, et al., 2010; Cummings, 2002).

In the last years, two different methods of obtaining Dye Sensitized Solar Cells have been studied (Mor, et al., 2006; Wang, et al., 2010; Flores, et al., 2007; Martinson, et al., 2007; Myahkostupov, et al., 2011), both aiming to develop new structures with improved properties in comparison with TiO_2 nanoparticles.

The arrangement of the highly ordered titania nanotube array perpendicular to the surface permits facile charge transfer along the length of the nanotube from the solution to the conductive substrate and simultaneously reduce the losses incurred by charge hopping across the nanoparticle grain boundaries (Gratzel, 2004; Law, et al., 2005; Tenne & Rao, 2004; Adachi, et al., 2003).

The enhancement in the electronic transport also allows the improved light harvesting as thicker films can be used to increase the optical density, thus improving the absorption of low-energy photons in the red and infrared without losing the additionally harvested charge carriers for recombination (Law, et al., 2005).

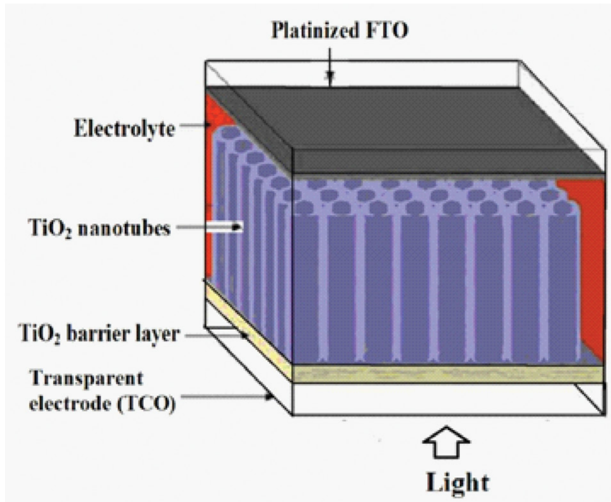
A highly-order TiO_2 nanotube dye-sensitized solar cell (Figure 9) consists of a TiO_2 nanotube electrode, coated with a suitable light-absorbing charge-transfer dye and wetted with an iodide/triiodide redox electrolyte. This compound is sandwiched between two Transparent and Conducting Oxide (TCO) glass. Usually fluorine doped SnO_2, exhibiting metallic conductivity, is used as the TCO. The TCO layer on the back glass substrate is covered with sputtered platinum, which acts as a catalyst for the redox reaction.

Theoretical Model

The simulation model (www.paper.edu.cn/index.php/default/feature/downCount/815) having a TiO_2 barrier layer, an electrolyte layer in TiO_2 nanotubes and a bulk electrolyte layer. This model introduces the TiO_2 barrier layer and the distinction between the electron transfer rate in the nanotube and in the nanoporous layer. The electrical model is based on that proposed by Ferber (Ferber, et al., 1998) and Oda (Oda, et al., 2006).

In the developed model (www.paper.edu.cn/index.php/default/feature/downCount/815), the electric field E has been ignored because of the negligible effect on charge carriers in the cell (Ferber, et al., 1998). A system of differential equations is obtained which is solved using the

Figure 9. Scheme of a cell structure with nanotube DSSC (Mor, et al., 2006)

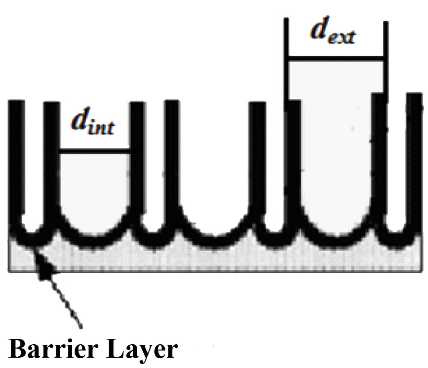


finite difference method (Ferber, et al., 1998). The equation system used in this model is very similar with the presented one. In this model, the ions n_{I^-}, $n_{I_3^-}$, and cations n_c depend on the porosity of TiO_2 nanotube layer. Figure 10 shows the view of nanotube arrays.

The porosity of nanotubes films is given by:

$$\rho_{nanotub} = 1 - \frac{V_e - V_i}{V} = 1 - \frac{\pi \left(r_{ext}^2 - r_{\text{int}}^2 \right) l}{3\sqrt{2}\ r_{ext}^2\ l} \tag{36}$$

Figure 10. Nanotube array with a corresponding top view



where

$r_{\text{int}} = \frac{d_{\text{int}}}{2}$, $r_{ext} = \frac{d_{\text{int}}}{2} + t_w$, dint = pore size, t_w = wall thickness, l = TiO_2 nanotube length

The incident photon flux density was calculated using the global distribution of the solar spectrum type AM 1.5 G, and has the value of $1000\ k\mathrm{W/m^2}$.

Material parameters are directly related to the performance parameters of the cell. The influence of the most important material parameters, such as electron mean lifetimes and mobilities, on the cell performance was investigated. The model permits the calculation of complete *J-V* curves, from which characteristical parameters of the solar cell are obtained: the short-circuit current (J_{sc}), the open-circuit voltage (V_{oc}), the fill factor (FF), and the energy conversion efficiency (η). Optimum parameters of Dye Sensitized Solar Cells based on TiO_2 nanotubes can be obtained.

Results and Discussion

The following values were obtained for characteristic parameters: $J_{sc} = 15.3\ \mathrm{mA/cm^2}$, $V_{oc} = 0.82$ V, $FF = 0.50$, and $\eta = 5.9\ \%$. These values are close to the experimental ones (Mor, et al., 2006; Wang, et al., 2010). In comparison with the present values obtained for nanocrystalline cells (Oda, et al., 2006): $J_{sc} = 16.1\ \mathrm{mA/cm^2}, V_{oc} = 0.81\ \mathrm{V}, FF = 0.72,$ and $\eta = 9.7\ \%$ the first ones are a little lower.

There are several reasons for the low efficiency of the nanotube DSSC cells:

- The used solar cells had the nanotubes of the negative electrode relatively short (of 6000 nm). This was reflected in a lower absorption than in the case of standard DSSC which have TiO_2 nanoparticles of 10 nm. It suggests the need of increasing the length of the nanotube array on the negative electrode as the amount of the absorbed dye appears to be the limiting factor.
- The low value of the form factor, $FF \sim 0.50,$ is caused by the serial resistance barrier introduced by the barrier layer thickness and poor contact between the barrier layer and FTO substrate. The annealing step used to crystallize the nanotube arrays can significantly increase the barrier layer thickness up to approximately 1 ¼m (Varghese, et al., 2003).
- The value of the conversion rate (power conversion efficiency) is influenced by the uniformity of the colorant absorbtion in the pores of the nanotube arrays. The nanotube has an array geometry which can absorb air and that is why the dye is not uniformly distributed.

CONCLUSION

Although the conversion rate of the nanotube solar cells is still low in this moment, a great potential can be foreseen.

REFERENCES

Adachi, M., Murata, Y., Okada, I., & Yoshikawa, S. (2003). Formation of titania nanotubes and applications for dye-sensitized solar cells. *Journal of the Electrochemical Society*, *150*, G488–G493. doi:10.1149/1.1589763

Adachi, M., Okada, I., Ngamsinlapasathian, S., Murata, Y., & Yoshikawa, S. (2002). Dye-sensitized solar cells using semiconductor thin film composed of titania nanotubes. *Electrochemistry*, *70*, 449–452.

Aernouts, T. (2006). *Organic bulk heterojunction solar cells: From single cell towards fully flexible photovoltaic module*. PhD Thesis. Belgium, The Netherlands: Katholieke Universiteit Leuven. Retrieved from https://lirias.kuleuven.be/bitstream/1979/402/5/PhDþThesisþTomþAernoutsþ18-09-06.pdf

Agnaldo, J. S., Cressoni, J. C., & Viswanathan, G. M. (2009). Universal photocurrent-voltage characteristics of dye sensitized nanocrystalline TiO2 photoelectrochemical cells. *Physical Review B: Condensed Matter and Materials Physics*, *79*(3), 035308. doi:10.1103/PhysRevB.79.035308

Berginc, M., Filipic, M., Krašovec, U. O., Nerat, M., Campa, A., & Hocevar, M. … Topic, M. (2009). *Electrical model of dye-sensitized solar cells*. Retrieved from http://www.solamon.eu/eng/Publications/Electrical-Model-of-Dye-Sensitized-Solar-Cells

Bowers, J. W., Upadhyaya, H. M., Nakada, T., & Tiwari, A. N. (2010). Effects of surface treatments on high mobility ITiO coated glass substrates for dye sensitized solar cells and their tandem solar cell applications. *Solar Energy Materials and Solar Cells*, *94*(4), 691–696. doi:10.1016/j.solmat.2009.10.023

Brabec, C., & Zerza, G., Cerullo, et al. (2001). Tracing photoinduced electron transfer process in conjugated polymer/fullerene bulk heterojunctions in real time. *Chemical Physics Letters*, *340*(3-4), 232–236. doi:10.1016/S0009-2614(01)00431-6

Cheng, H. M., & Hsieh, W. F. (2010). High-efficiency metal-free organic-dye-sensitized solar cells with hierarchical ZnO photoelectrode. *Energy and Environmental Science*, *3*(4), 442–447. doi:10.1039/b915725e

Chiba, Y., Islam, A., Watanabe, Y., Komiya, R., Koide, N., & Han, L. (2006). dye-sensitized solar cells with conversion efficiency of 11.1%. *Japanese Journal of Applied Physics*, *45*, L638–640. doi:10.1143/JJAP.45.L638

Chuan, J., Tianze, L., Luan, H., & Xia, Z. (2011). Research on the characteristics of organic solar cells. *Journal of Physics Conference Series*, *276*(1), 012169. doi:10.1088/1742-6596/276/1/012169

Chuan, J., Tianze, L., Luan, H., & Xia, Z. (2011). Research on the characteristics of organic solar cells. *Journal of Physics: Conference Series*, *276*(1).

Contreras, M. A., Egaas, B., Ramanathan, K., Hiltner, J., Swartzlander, A., Hasoon, F., & Noufi, R. (1999). Progress toward 20% efficiency in Cu(In,Ga)Se2 polycrystalline thin-film solar cells. *Progress in Photovoltaics: Research and Applications*, *7*(4), 311–316. doi:10.1002/(SICI)1099-159X(199907/08)7:4<311::AID-PIP274>3.0.CO;2-G

Cummings, F. (2010). *Application of TiO_2 nanotubes in dye-sensitised solar cells for improved charge transport.* Paper presented at the Renewable Energy Photovoltaics, SA Germany Workshop on Photovoltaics. Grahamstown, Germany.

Deb, K. S. (2004). Dye-sensitized TiO_2 thin-film solar cell research at the national renewable energy laboratory (NREL). *Solar Energy Materials and Solar Cells*, *88*(1), 1–10. doi:10.1016/j.solmat.2004.09.007

Dennler, G., Mozer, A. J., Juška, G., Pivrikas, A., Österbacka, R., Fuchsbauer, A., & Sariciftci, N. S. (2006). Charge carrier mobility and lifetime versus composition of conjugated polymer/fullerene bulk-heterojunction solar cells. *Organic Electronics*, *7*, 229–234. doi:10.1016/j.orgel.2006.02.004

Dimitrakopoulos, C. D., & Mascaro, D. J. (2001). Organic thin-film transistors: A review of recent advances. *IBM Journal of Research and Development*, *45*, 11–28. doi:10.1147/rd.451.0011

Ferber, J., & Luther, J. (2001). Modeling of photovoltage and photocurrent in dye-sensitized titanium dioxide solar cells. *The Journal of Physical Chemistry B*, *105*(21), 4895–4903. doi:10.1021/jp002928j

Ferber, J., Stangl, R., & Luther, J. (1998). An electrical model of the dye-sensitized solar cell. *Solar Energy Materials and Solar Cells*, *53*(1-2), 29–54. doi:10.1016/S0927-0248(98)00005-1

Flores, I. C., de Freitas, J. N., Longoa, C., De Paoli, M., Winnischofer, H., & Nogueira, A. F. (2007). Dye-sensitized solar cells based on TiO2 nanotubes and a solid-state electrolyte. *Journal of Photochemistry and Photobiology A Chemistry*, *189*, 153–160. doi:10.1016/j.jphotochem.2007.01.023

Gao, F., Wang, Y., Shi, D., Zhang, J., Wang, M. K., & Jing, X. Y. (2008). Enhance the optical absorptivity of nanocrystalline TiO_2 film with high molar extinction coefficient ruthenium sensitizers for high performance dye-sensitized solar cells. *Journal of the American Chemical Society*, *130*(32), 10720–10728. doi:10.1021/ja801942j

Giannouli, M., Syrrokostas, G., & Yianoulis, P. (2010). Effects of using multi-component electrolytes on the stability and properties of solar cells sensitized with simple organic dyes. *Progress in Photovoltaics: Research and Applications*, *18*(2), 128–136. doi:10.1002/pip.952

Grätzel, M. (2003). Dye-sensitized solar cells. *Journal of Photochemistry and Photobiology C, Photochemistry Reviews*, *4*(2), 145–153. doi:10.1016/S1389-5567(03)00026-1

Gratzel, M. (2004). Conversion of sunlight to electric power by nanocrystalline dye-sensitized solar cells. *Journal of Photochemistry and Photobiology A Chemistry*, *168*(1-3), 235–237. doi:10.1016/j.jphotochem.2004.08.014

Grätzel, M. (2006). The advent of mesoscopic injection solar cells. *Progress in Photovoltaics: Research and Applications*, *14*(5), 429. doi:10.1002/pip.712

Green, M. A., Emery, K., Hishikawa, Y., & Warta, W. (2011). Solar cell efficiency tables (version 37). *Progress in Photovoltaics: Research and Applications*, *19*(1), 84–92. doi:10.1002/pip.1088

Greulich, J., Markus Glatthaar, M., & Rein, S. (2010). Fill factor analysis of solar cells' current–voltage curves. *Progress in Photovoltaics: Research and Applications*, *18*(7), 511–515. doi:10.1002/pip.979

Hardin, B. E., Hoke, E. T., Armstrong, P. B., Yum, J. H., Comte, P., & Torres, T. (2009). Increased light harvesting in dye-sensitized solar cells with energy relay dyes. *Nature Photonics*, *3*(11), 667. doi:10.1038/nphoton.2009.212

Hiramoto, M., Fujiwara, H., & Yokoyama, M. (1991). Three-layered organic solar cell with a photoactive interlayer of codeposited pigments. *Applied Physics Letters*, *58*(10), 1062–1064. doi:10.1063/1.104423

King, R. R., Law, D. C., Edmondson, K. M., Fetzer, C. M., Kinsey, G. S., & Yoon, H. (2007). 40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells. *Applied Physics Letters*, *90*(18), 183516–1835111. doi:10.1063/1.2734507

Knodler, R., Sopka, J., Harbach, F., & Grunling, H. W. (1993). Photoelectrochemical cells based on dye sensitized colloidal TiO_2 layers. *Solar Energy Materials and Solar Cells*, *30*(3), 277–281. doi:10.1016/0927-0248(93)90147-U

Kwok, H. L. (2003). Carrier mobility in organic semiconductor thin films. *Reviews on Advanced Materials Science*, *5*, 62–66.

Lai, Y. H., Lin, C. Y., Chen, J. G., Wang, C. C., Huang, K. C., & Liu, K. Y. (2010). Enhancing the performance of dye-sensitized solar cells by incorporating nanomica in gel electrolytes. *Solar Energy Materials and Solar Cells*, *94*(4), 668–674. doi:10.1016/j.solmat.2009.11.027

Law, M., Greene, L. E., Johnson, J. C., Saykally, R., & Yang, P. D. (2005). Nanowire dye-sensitized solar cells. *Nature Materials*, *4*, 455–459. doi:10.1038/nmat1387

Marlein, J., & Burgelman, M. (2007). Electrical properties of CIGS cells. In *Proceedings of the 22nd European Photovoltaic Solar Energy Conference*. Milan, Italy: EU PVTSEC.

Marlein, J., Decock, K., & Burgelman, M. (2009). Analysis of electrical properties of CIGSSe and Cd-free buffer CIGSSe solar cells. *Thin Solid Films*, *517*(7), 2353–2356. doi:10.1016/j.tsf.2008.11.048

Martinson, A. B. F., Elam, J. W., Hupp, J. T., & Pellin, M. J. (2007). ZnO nanotube based dye-sensitized solar cells. *Nano Letters*, *7*(8), 2183–2187. doi:10.1021/nl070160+

Meng, Q.-B., Takahashi, K., Zhang, X.-T., Sutanto, I., Rao, T. N., Sato, O., & Fujishima, A. (2003). Fabrication of an efficient solid-state dye-sensitized solar cell. *Langmuir*, *19*, 3572. doi:10.1021/la026832n

Mitroi, M. R., Iancu, V., Fara, L., & Ciurea, M. L. (2011). Numerical analysis of J–V characteristics of a polymer solar cell. *Progress in Photovoltaics: Research and Applications*, *19*(3), 301–306. doi:10.1002/pip.1026

Mor, G. K., Varghese, O. K., Paulose, M., Shankar, K., & Grimes, C. A. (2006). A review on highly ordered, vertically oriented TiO2 nanotube arrays: Fabrication, material properties, and solar energy applications. *Solar Energy Materials and Solar Cells*, *90*, 2011–2075. doi:10.1016/j.solmat.2006.04.007

Myahkostupov, M., Zamkov, M., & Castellano, F. N. (2011). Dye-sensitized photovoltaic properties of hydrothermally prepared TiO2 nanotubes. *Energy and Environmental Science*, *4*, 998–1010. doi:10.1039/c0ee00533a

Nazeeruddin, M. K., De Angelis, F., Fantacci, S., Selloni, A., Viscardi, G., & Liska, P. (2005). Combined experimental and DFT-TDDFT computational study of photoelectrochemical cell ruthenium sensitizers. *Journal of the American Chemical Society*, *127*(48), 16835–16847. doi:10.1021/ja052467l

Nelson, J. (2003). Diffusion-limited recombination in polymer-fullerene blends and its influence on photocurrent collection. *Physical Review B: Condensed Matter and Materials Physics*, *67*, 155209. doi:10.1103/PhysRevB.67.155209

Nunzi, J. M. (2002). Organic photovoltaic materials and devices. *Comptes Rendus Physique*, *3*(4), 523–542. doi:10.1016/S1631-0705(02)01335-X

O'Regan, B., & Grätzel, M. (1991). A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films. *Nature*, *353*, 737–740. doi:10.1038/353737a0

Oda, T., Tanaka, S., & Hayase, S. (2006). Differences in characteristics of dye-sensitized solar cells containing acetonitrile and ionic liquid-based electrolytes studied using a novel model. *Solar Energy Materials and Solar Cells*, *54*(16), 2696–27011. doi:10.1016/j.solmat.2005.11.013

Paper. (2011). *Website.* Retrieved from http://www.paper.edu.cn/index.php/default/feature/downCount/815

Shaheen, S., Brabec, C. J., Sariciftci, N. S., Padinger, F., Fromherz, T., & Hummelen, J. C. (2001). 2.5% efficient organic plastic solar cells. *Applied Physics Letters*, *78*, 841–843. doi:10.1063/1.1345834

Smestad, G., Bignozzi, C., & Argazzi, R. (1994). Testing of dye sensitized TiO2 solar cells I: Experimental photocurrent output and conversion efficiencies. *Solar Energy Materials and Solar Cells*, *32*(3), 259–272. doi:10.1016/0927-0248(94)90263-1

Soedergren, S., Hagfeldt, A., Olsson, J., & Lindquist, S. E. (1994). Theoretical models for the action spectrum and the current-voltage characteristics of microporous semiconductor films in photoelectrochemical cells. *Journal of Physical Chemistry*, *98*(21), 5552–5556. doi:10.1021/j100072a023

Tang, C. W. (1986). Two-layer organic photovoltaic cell. *Applied Physics Letters*, *48*(2), 183–185. doi:10.1063/1.96937

Tenne, R., Rao, C. N. R. (2004). Inorganic nanotubes. *Philosophical Transactions of the Royal Society A, 362*(1823), 2099-2125.

Tian, H., Zhang, J., & Yangjing, Yu. T., & Zou, Z. (2011). Analysis on dye-sensitized solar cell's efficiency improvement. *Journal of Physics: Conference Series, 276*, 012188.

Varghese, O. K., Gong, D., Paulose, M., Ong, K. G., Dickey, E. C., & Grimes, C. A. (2003). Extreme changes in the electrical: Resistance of titania nanotubes with hydrogen exposure. *Advanced Materials (Deerfield Beach, Fla.)*, *15*, 624–627. doi:10.1002/adma.200304586

Wang, Y., Yang, H., Liu, Y., Wang, H., Shen, H., Yan, J., & Xu, H. (2010). The use of Ti meshes with self-organized TiO_2 nanotubes as photoanodes of all-Ti dye-sensitized solar cells. *Progress in Photovoltaics: Research and Applications*, *18*, 285–290.

Wei, D. (2010). Dye sensitized solar cells. *International Journal of Molecular Sciences*, *11*, 1103–1113. doi:10.3390/ijms11031103

Wenger, S. (2010). *Strategies to optimizing dye-sensitized solar cells: Organic sensitizers, tandem device structures, and numerical device modeling*. Retrieved from http://library.epfl.ch/en/theses/?nr=4805

Wenger, S., Schmid, M., Rothenberger, G., Grätzel, M., & Schumacher, J. O. (2009). *Model-based optical and electrical* characterization of dye-sensitized solar cells. Retrieved from http://www.juergenschumacher.de/html/material/veroeffentlichungen/DSC-EUPVSEC-Hamburg.pdf

Wienk, M., Kroon, J. M., Verhees, W. J. H., Krol, J., Hummelen, J. C., Van Haal, P., & Janssen, R. A. J. (2003). Efficient methano[70] fullerene/MDMO-PPV bulk heterojunction photovoltaic cells. *Angewandt Chemie International Edition, 42*(29), 3371-3375.

Xue, J., Rand, B. P., Uchida, S., & Forrest, S. R. (2005). A hybrid planar-mixed molecular heterojunction photovoltaic cell. *Advanced Materials (Deerfield Beach, Fla.)*, *17*(1), 66–71. doi:10.1002/adma.200400617

Yong, V., Ho, S. T., & Chang, R. P. H. (2008). Modeling and simulation for dye-sensitized solar cells. *Applied Physics Letters*, *92*(14), 143506. doi:10.1063/1.2908204

Zhang, X.-T., Liu, H.-W., Taguchi, T., Meng, Q. B., Sato, O., & Fujishima, A. (2004). Slow interfacial charge recombination in solid-state dye-sensitized solar cell using Al_2O_3-coated nanoporous TiO_2 films. *Solar Energy Materials and Solar Cells*, *81*(2), 197–203. doi:10.1016/j.solmat.2003.11.005

Zhao, J., Wang, A., Green, M. A., & Ferrazza, F. (1998). 11.8% efficient "honeycomb" textured multicrystalline and 24.4% monocrystalline silicon solar cells. *Applied Physics Letters*, *73*(14), 1991–1993. doi:10.1063/1.122345

This work was previously published in Advanced Solar Cell Materials, Technology, Modeling, and Simulation, edited by Laurentiu Fara and Masafumi Yamaguchi, pp. 120-137, copyright 2013 by Engineering Science Reference (an imprint of IGI Global).

Chapter 105
Programming Robots in Kindergarten to Express Identity: An Ethnographic Analysis

Marina Umaschi Bers
Tufts University, USA

Alyssa B. Ettinger
Tufts University, USA

ABSTRACT

This chapter presents a research program that uses robotics as a powerful tool to engage Kindergarten children in developing computational thinking and learning about the engineering design process. Using an ethnographic analysis of an experience in a Kindergarten classroom at the Jewish Community Day School (JCDS), a pluralistic school in Watertown, MA, in which children worked with robotics as a way to explore issues of identity, the chapter highlights both developmental and technological considerations that need to be addressed when engaging young children with robotic activities. This project used an innovative hybrid tangible programming system composed of interlocking wooden blocks, called CHERP, specifically designed to meet the developmental needs of young children. While many robotic programs highlight building aspects and their relationship to engineering education, the approach presented in this chapter complements this by focusing on programming by teaching powerful ideas from computer science at a very early age.

DOI: 10.4018/978-1-4666-1945-6.ch105

INTRODUCTION

Typically, "robotics" brings to mind metallic human-like contraptions wired with complex electronics. However, this chapter describes an experience in which simple Lego-based robotic cars were programmed by Kindergarten children with smart wooden blocks using CHERP (Creative Hybrid Environment for Robotic Programming), a developmentally appropriate tangible language (Horn, Crouser & Bers, 2011). This work was inspired by the realization that in the early grades, children learn very little about engineering and technology. Just as it is important to begin science instruction in the early years by building on children's curiosity about the natural world, it is as important to begin engineering instruction and the development of technological literacy by building on children's natural inclination to design and build things, and to take things apart to see how they work (Bers, 2008; Petroski, 2003). Robotics is a wonderful platform that taps into what is unique to today's human-made world: the fusion of electronics with mechanical structures.

With the growing popularity of robotics, the use of educational robotic kits and programming languages for controlling the robot's behaviors is becoming widespread in high schools, middle and elementary schools (Rogers, Wendell & Foster, 2010). In order to bring robots to "life" children must create computer programs—digital artifacts that allow robots to move, blink, sing, and respond to their environment. Previous research has shown that children as young as four years old can understand the basic concepts of computer programming and can build and program simple robotics projects (Bers, 2008; Cejka, Rogers, & Portsmore, 2006; Bers et al, 2006; Bers & Horn, 2010; Kazakoff & Bers, 2010; Bers, 2010a). However, young children need to work with interfaces that are developmentally appropriate. The robotics-based programming language we used, called CHERP, is such a tool and was developed by Bers and her DevTech research team at Tufts University (Horn et al., 2011). Rather than writing computer programs with a keyboard or mouse, the CHERP system allows children to instead *construct* physical computer programs by connecting interlocking wooden blocks. CHERP is described in the following section.

This chapter takes an ethnographic approach to analyze and describe the learning experience of 23 Kindergarten students who participated in a month long robotics curriculum called TangibleK, developed by the DevTech research group at Tufts University with funding from the National Science Foundation. The TangibleK curriculum, which utilizes CHERP to teach robotics and computer programming concepts to Kindergarten students, was adapted and extended to explore issues of identity at the pluralistic Jewish Community Day School (JCDS) in Watertown, MA. While the TangibleK curriculum encourages cognitive development in such areas as logical and sequential thinking (Kazakoff & Bers, 2010), the overarching project goal at JCDS was not only to engage children in learning about robotics, but also to provide them with robotics as a different medium, to express their explorations of identity in a creative way.

We collaborated with JCDS Kindergarten teachers to incorporate the TangibleK curriculum into the class' end-of-year project that encouraged the students' reflection of their accomplishments during the school year and their developing sense of self. This culminating project was called *Mi Ani* ("Who am I," in Hebrew). The extension to the TangibleK curriculum for the *Mi Ani* project focused on the creation of robotic artifacts and programmed behaviors to express the kindergartners' individual Jewish identities. Because the medium of robotics allows the display of actions, as opposed to static facts, children chose to create robots enacting behaviors that are related to their different ways of being Jewish. For example, one student programmed his robot to spin to represent lighting the Hanukkah menorah, while another programmed hers to roll back and forth, mimicking rolling out dough for Passover matzah. The robots created by the children represented a sense of dynamic identity, "chosen" by the children. The dynamic process that enhanced the conceptualization of identity involved engaging children in deciding how to make their robots, what aspects of their Jewish identities to convey through them, and how to program them to respond to certain events in the Jewish calendar (Libman, 2011; Bers & Urrea, 2000).

CONTEXT OF STUDY

Setting

At JCDS, 23 children in the *Gan Nitzan (Kindergarten)* class participated in the Mi Ani project using the TangibleK robotics curriculum developed by the DevTech team at Tufts University (Bers, 2010b). The philosophy of the school is one that supports progressive, child-centered learning styles, which provided a welcoming environment for the TangibleK curriculum. The school is philosophically committed to intentional pluralism, meaning that it actively embraces children and families with a wide range of Jewish expression, practice, and belief. Pluralism at the school extends beyond variations of Jewish religious affiliation. For example, the school's curriculum embraces differentiated instruction, celebrating the diversity of learners represented within the school's community.

Gan Nitzan is a combination of the Hebrew words *gan*, meaning 'Kindergarten' or 'preschool' and *nitzan,* meaning 'flower bud'. The Kindergarten has full days of school Monday through Thursday (8:00 am to 3:30 am) and a shorter day on Fridays (ending at 2:00 pm) in anticipation of *Shabbat*, the Jewish Sabbath. The Kindergarteners have two main teachers, and often an assistant teacher as well, to achieve a low student-teacher ratio. The *Gan Nitzan* classroom balances a developmentally sensitive academic structure with an emphasis on supporting the Kindergarteners' creativity and individual learning styles. The main curriculum integrates academic studies with Judaic studies, and fosters respect for multiple perspectives while promoting curiosity and a passion for learning. The main curriculum comprises English literacy studies, integrated Hebrew lessons, *Humash* (Bible) study, math, science, and social studies. Specialists also teach several additional subjects: physical education, art, music, Israeli dancing, and movement (a class that promotes body awareness). Of note, each grade at the JCDS has an overarching "central subject" that shapes all main areas of their curriculum. The Kindergarten's central subject is "Cycles", which is explored across multiple domains, such as the study of the cyclical Jewish holidays, the life cycle of plants and butterflies, and recognition of patterns in stories. *Gan Nitzan's* final project, *Mi Ani*, highlights the cyclical nature of time while celebrating the self-reflection of the Kindergarteners' experiences throughout the school year.

The *Mi Ani* project lasted for a period of one month and culminated in a final open house for family and friends where children presented their robots. One of the two main classroom teachers, who was trained in the technology by the DevTech Team, and two DevTech researchers, were in charge of classroom instruction. The curriculum that formed the foundation of the *Mi Ani* project was implemented in two stages. First, children were introduced to the robotic technology and the CHERP programming language by engaging with the TangibleK curriculum. This curriculum taught children computer programming and engineering concepts and provided a structured way for children to put those concepts to use by engaging in small-scale activities and challenges. Once children mastered the basics of programming and building robots, the second part of the curriculum focused more specifically on the *Mi Ani* project. Together with their teachers, children reflected on their experiences and learning during the year, coming up with a timeline consisting of the differ-

Figure 1. A section of the timeline of the Kindergarten year

ent activities, during each academic month, that were meaningful to them (see Figure 1).

Each child chose three moments in the year as "stations" at which his or her robot would stop and perform a program to represent the child at that given moment. For example, one child programmed his robot to stop along the timeline at November, spinning to represent eating turkey on Thanksgiving, while another programmed her robot to sing at December to show singing Hanukkah songs. The children decorated the robots to represent their self-identities, using art materials to depict them, their interests, and their characteristics. For example, one child decorated her robot with drawings of all her favorite sports, while another molded a clay image of herself that she attached to the top of her robot. Finally, children combined their three robot programs from individual moments in the year into one comprehensive program, which represented their journey throughout the year. Each child programmed his or her robot to travel alongside the timeline, demonstrating the children's own understanding of significant moments of their experience throughout the year.

CHERP in the Classroom

The teachers created an environment that emphasized an engineering mindset, with a focus on teaching the Engineering Design Process and with the impetus to engage Kindergarteners to consciously take on the identity of engineers. Students were immersed in a community of engineers that shared a set of core values, knowledge, skills, and identities. At the introduction of the TangibleK curriculum, the Kindergarteners were explicitly introduced to the Engineer Design Process, which was presented to the class as a visual web that cycled through the five steps: *ask, imagine, plan, create, improve*. The teachers discussed and provided examples for each step, emphasizing that the students would become engineers when they engaged in the process of designing and programming their own robots.

After about one week of lessons on sturdy building with Lego blocks and understanding what a robot is, the children were actively taught the robotics technology. The TangibleK program uses CHERP in combination with different robotic kits. For this project we used the Lego MINDSTORMS® RCX brick, an embedded micro-computer or the "robot brain". The programs children created with the CHERP tangible blocks were downloaded to the Lego RCX brick using infrared. The CHERP programming language draws on work from the field of human computer interaction (HCI) on tangible interfaces that shows that we can overcome the inherent limitations of writing computer programs with a keyboard or mouse, by offering tangible systems (Blikstein, Buechley, Horn & Raffle, 2010). Tangible languages, instead of relying on pictures and words on a computer screen, use physical objects to represent the various aspects of computer programming. Users arrange and connect these physical elements to construct programs. Tangible languages exploit the physical properties of objects such as size, shape, and materials to express and enforce syntax. This is crucial when working with Kindergarten age children who might not have developed the motor skills to work with a mouse interface or who might need physical objects to manipulate in order to

Figure 2. The CHERP language developed at Tufts University. Children construct programs for their robots using interlocking wooden blocks.

understand abstract concepts (Horn et al, 2011; Bers & Horn, 2010).

With the CHERP system, young children can transition back and forth between using interlocking tangible wooden blocks, or onscreen programs using the same icons that represent actions for their robots to perform (see Figure 2). This hybrid approach allows children to work with multiple representations (Horn et al, 2011).

CHERP uses a collection of image processing techniques to convert physical programs into digital instructions. A standard webcam connected to a desktop or laptop computer takes a picture of the program. A compiler converts the picture into digital code that gets downloaded to the robot in a few seconds.

Children worked in eleven pairs, while one child worked alone. This set-up fostered the potential for collaboration and social development along with cognitive gains. Each group shared an RCX brick to build a car, onto which they connected two Lego motors and wheels and a "leg" in the front to keep the base upright. Each pair also had a mini laptop station set up with the CHERP software, a webcam to scan in the wooden block, and a Lego tower to upload the programs to the RCX brick. Large bins of Legos and CHERP wooden blocks were available on the rug to be shared by all of the students.

TangibleK Curriculum: Introduction to Programming

The overarching goal of the TangibleK curriculum is to introduce young children to computational thinking. Computational thinking is a type of analytical thinking that shares many similarities with mathematical thinking (e.g., problem solving), engineering thinking (designing and evaluating processes), and scientific thinking (systematic analysis). The foundation for computational thinking is abstraction – abstracting concepts from cases and evaluating and selecting the "right" abstraction. It relies on selection of inputs (manipulation of variables and computational instructions), observation of outputs (outcome data), and decomposition of what happens in between. Computational thinking is about the ability to abstract from computational instructions (programming languages) to computational behaviors, to identify potential "bugs" and places for errors, to decide what details among the input-computation-output algorithm to highlight and retain and what details to discard (Wing, 2006).

The term computational thinking grew out of the pioneer work of Seymour Papert and colleagues on design-based constructionist programming environments, to refer to ways to algorithmically solve problems and to the acquisition of technological fluency (Papert, 1980; Papert, 1993). Previous work on young elementary school-aged children and computational thinking can be found in the research literature on constructionist programming environments (Repenning, Webb, & Ioannidou, 2010; Resnick, Maloney, Monroy-Hernandez, Rusk, Eastmond, Brennan, et al., 2009, 2009). Wing (2006) describes computational thinking as a fundamental skill for everyone, not just for computer scientists.

In order to engage children in computational thinking, the TangibleK curriculum focuses on the

Table 1. The six powerful ideas of the TangibleK program

Powerful idea	*Definition*	*Activity*	*Discipline connections*
Robotics	The engineering discipline that focuses on the creation and programming of robots, machines that can follow instructions and move on their own to perform tasks.	*What Is a Robot?* After an introduction to robotics by looking at different robots and talking about the functions they serve, children build their own robotic vehicles and explore their parts and the instructions they can use to program them.	• Engineering • Computer Science
Engineering design process	A process used to develop products to solve a need or problem. It has several iterative steps: identifying a need or defining the problem, doing research, analyzing possible solutions, developing the product, communicating and presenting the product.	*Sturdy building:* Children build a non-robotic vehicle to take small toy people from home to school. The vehicle needs to be sturdy as well as perform its intended functions.	• Engineering • Computer science
Sequencing / control flow	A sequence of instructions can be described in a program and acted out in order by a robot. Each block has a specific meaning. The order of the blocks is important.	*The Hokey-Pokey:* Choose the appropriate commands and put them in order to program a robot to dance the Hokey-Pokey.	• Creative storytelling • Organization of ideas • Mathematical proofs • Procedural thinking
Loops and Parameters	A sequence of instructions can be modified to occur over and over again. Control flow commands can be qualified with additional information. For example, loops can be modified to repeat forever or a concrete number of times.	*Again and Again until I Say When:* Students use a pair of loop blocks ("repeat" / "end repeat") to make the robot go forward again and again, infinitely and then just the right number of times to arrive at a fixed location.	• Cyclical events in nature • Scheduling • Timing and control • Feedback loops • Number sense
Sensors	A robot can use sensors, akin to human sense organs, to gather information from its environment. Sensor information can be used to control when the robot follows given commands.	*Through the Tunnel:* Children use light sensors and commands to program a robot to turn its lights on when its surroundings are dark and vice versa.	• Scientific observations • Cause and effect • Sensors (both human-made and natural)
Branches	At a branch in the program, a robot can follow one set of commands or another depending on the state of a given condition.	*The Robot Decides:* Students program their robot to travel to one of two destinations based on light or touch sensor information.	• Cause and effect • Sensors (both human-made and natural)

following powerful ideas: robotics, engineering design process, sequencing and control flow, loops and parameters, sensors and branches. See Table 1 for descriptions of these.

The TangibleK curriculum is designed for a minimum of 20 hours of classroom work, divided into the following structured sessions based on the six powerful ideas identified above:

1) Sturdy building (the engineering design process);
2) What is a robot? (robots have special parts to follow instruction);
3) Hokey-Pokey: sequence of commands (the sequence or order of commands matters);
4) Again & Again Until I Say When (loops and number parameters);
5) Through the tunnel (sensors and loops) ; and
6) The robot decides (sensors and branches).

In the first phase of the curriculum, the teachers taught the CHERP system one lesson at a time, with directed challenges following the teaching of each new type of block representing an action. One of the first lessons challenged the children to make their robots dance the "Hokey Pokey". This required mastery of the CHERP syntax and the sequential nature of computer programming. For example, the "Hokey Pokey" program used the following sequence of commands:

"BEGIN/FORWARD/BACKWARD/FORWARD/SHAKE/SPIN/END".

The robotics terms were taught by comparing electronic parts to human body parts, such as the "ear" of the robot that receives the signals of the program from the computer tower. During one morning meeting, a teacher led a very successful game where a student acted out his own program and the class had to guess what the program blocks were. Bodrova and Leong (2007) emphasize play as the time when most learning takes place, in particular during cooperative dramatic play, which involves taking on different roles. Activities in which the children and robots took on each other's roles likely allowed the Kindergarteners to better learn how the robots work through a sort of dramatic play.

ETHNOGRAPHIC ANALYSIS OF ROBOTICS EXPERIENCE

Our methodology for this ethnographic study involved the full immersion of the ethnographer into the *Gan Nitzan* classroom during the month of the TangibleK curriculum. The ethnographer was introduced to the class as an assisting member of the DevTech research team, and was familiar with the TangibleK technology and software so that she could actively participate in the curriculum. She worked with the students throughout the curriculum and conducted face-to-face interviews with the children as they developed their robotics skills and projects. Observation notes were compiled each day along with frequent audiovisual documentation, and the ethnographer helped administer the post-curriculum assessments of the students at the

Figure 3. The CHERP program for performing the "Hokey Pokey" dance, as it would appear on the graphical computer interface

end of the curriculum. During robotics lessons, the ethnographer was free to navigate the classroom to observe, interview, and offer help to those students who needed it. The role of the ethnographer as an active participant in the classroom allowed her to form positive relationships with the children and to experience each phase of the curriculum from their perspectives.

Programming Challenges and Learning to Fail

As the children were introduced to complex blocks, the programming challenges became more difficult and children faced confusion and frustration as they began to deal with many more failures than successes. For example, one of the later lessons exposed the children to light sensors and light bulbs that can connect to the robot brain. One of the head JCDS teachers began the lesson asking the children to think about how we would tell whether it is day (light) or night (dark), highlighting the comparison between the robots' light sensors and our eyes. This attribution of human characteristics to the robot parts again allowed the children to easily grasp the concept of the light sensors. Utilizing this new material and building upon the repeat loops from the previous lesson, the challenge of the day was to create a program that took their robot down a straight "street" and through a dark tunnel, turning on its light when it was inside the tunnel, and then back off when it reached the other side. This challenge was certainly more complex than previous ones because it necessitated repeat loops with light sensors as the parameter, a novel combination that was not explicitly modeled for the children.

Perhaps because of the steady pace at which the curriculum was moving, most children were not yet comfortable enough with the repeat loops to manipulate them for such an abstract challenge. Without being able to work from a model, many children began the challenge with little direction and were easily distracted by their classmates. Subsequent prompting from the teachers did provide some scaffolding for a few students to come close to solving the challenge, but for the first time the majority of students were not successful. Many children were still unsure of how exactly to use the repeat loops, especially with qualitative rather than quantitative parameters.

As a result, the class did not necessarily master the conceptual content, but the lesson was crucial in establishing the mindset that initial failure is acceptable. The students were always encouraged to continuously test out their robot and modify their program, following the Engineer Design Process, but this was one of the first challenges in which there were multiple solutions and many more opportunities for trial and error. A compounding difficulty was that the light sensors themselves proved finicky and resulted in a false negative feedback for most of the students when they tested their robot's performance of the programs.

During the last few minutes of the robotics period, the teachers led "Technology Circles" for the students to share their programs and talk about their successes and difficulties. The teachers made sure to emphasize that it was okay to have to keep on trying and that even our robots did not work perfectly with the light sensors. Remarkably, a few pairs did create working programs by the end of the day, and the rest at least learned a valuable lesson in emotional regulation and perseverance.

Mastering Concepts of Computer Programming: Earning "License"

Once the children had learned the robotics content, the second phase of the curriculum was a final project for children to explore how to express, through their robots, a sense of Jewish identity. The Mi Ani (Who am I?) project allowed the children to use their knowledge of robotics creatively to design and program personalized robots representing their own journeys through the school year. Each individual robot would traverse a linear timeline of large "month" posters filled with photographs of the students from each month of the year and would stop at three months of the child's choice. At each point, the robot would perform a series of actions that corresponded to the child's interpretation of a memorable school event that month. Here the students were given an open-ended opportunity to express their identities and experiences, and were encouraged to be creative, playful, and imaginative. In some ways, the thought process shifted here from the traditional Engineer Design Process to the Creative Thinking Spiral (imagine, create, play, share, reflect), which emphasizes imagination and reflection over solving a problem (Resnick, 2007).

Before the Kindergarteners began their final projects, they were required to demonstrate essential robotics and programming tasks to a teacher in order to earn their "licenses". Since we had observed that many students still had not grasped some of the robotics and programming concepts or relied on their partner for programming their robot, the teachers decided to use the license as an individual assessment method and right of passage to being able to progress to the Mi Ani projects. The students received stickers on notecards for successfully completing such tasks as building a sturdy robot, connecting wires correctly, creating a working program, and uploading the program to the robot. Having concrete tasks presented step-by-step and with immediate gratification dramatically improved the Kindergarteners' motivation and performance. In contrast to most other lessons, almost every student remained on task until she had completed her license, and most partners were able to negotiate taking turns with the robot, even when they had not worked well together in the past.

The licenses proved to be a wonderful assessment tool to ensure that everyone was at least at a baseline understanding of the robotics and to boost the students' confidence and motivation with a sense of personal achievement and mastery. It is worth noting that at the beginning of the curriculum, most students were drawn to the blocks as their choice of interface, but by the time they arrived at the license phase, every student had switched over to the graphical interface to complete the tasks. Perhaps the tangible blocks was the appropriate interface to initially transition the Kindergarteners to robotics because the coded blocks closely resembled familiar toys they were used to playing with. However, once the students were comfortable enough with the abstract concepts of programming, they likely favored the graphical interface because of its convenience. This switch is significant because it suggests that the Kindergarteners had achieved a sense of symbolic representation – they could view the 2-dimensional squares on the computer screen as abstract representations of their 3-dimensional tangible blocks.

Mi Ani?: Personalization and Mindful Planning

Once the children were cleared to start their Mi Ani projects, they were shown how to change the "channel", or program number, on the RCX so that each partner could create and upload his separate program onto different memory slots in the same robot. To further provide the students with a sense of identity and ownership for their robots, each child created her own platform reflecting her identity, which she would secure to the robot when running her program. The students were provided with a wide range of arts and crafts materials, including Plasticine modeling clay, paper, markers, colored wires, and Lego with which to design their platforms. The platforms provided an opportunity to combine robotics with traditional arts and crafts materials and were also key to the children's sense of ownership and personal investment in their projects. The children utilized diverse designs and materials, with images ranging from favorite foods, animals and colors to symbols of their religious identity, such as Jewish stars. Many students also drew JCDS signs, demonstrating a strong sense of community and close identification with their school.

The children brainstormed significant events and activities from each month of the school year with their head teachers, looking at photographs that the teachers had taken to document their experiences. Each child eventually chose three favorite events from different months and filled out her own design journal to plan out their robot's program by gluing down paper cut-outs of the blocks. The design journals also contained the child's explanation of the chosen event and the meaning behind the robot's actions. Although most of the students' programs inevitably changed from the original plan, the design journals were extremely helpful in making the programming stage more efficient because students had already formulated ideas before sitting down at the computer. The guiding questions in the journal also emphasized the reflective nature of the project, helping the students remember that their programs needed to be thoughtfully and meaningfully planned.

For the week leading up to the Mi Ani final presentation, students moved between several stations: choosing photos to decorate posters for the months, working on design journals, and programming their latest event. As the children began programming their robots, most of them opted for simplicity until they were urged to revise or modify their programs by a teacher. For example, students would put a series of five forwards in a row instead of using only three blocks in a repeat loop to achieve the same motion (REPEAT/5/FORWARD). Once prompted by a teacher or researcher, most of the children demonstrated sufficient knowledge of how to utilize the repeat blocks.

The personal investment that children had in their Mi Ani projects motivated each student to put more effort into her own robot and program than they had for previous lessons with their partners. Although many students noticeably developed more self-regulation and patience in the face of obstacles over the course of the curriculum, their personal interest in their project's success sometimes made failures more devastating.

Student Achievements in Expressing Their Identity through Robotics

Overall, the students came up with thoughtful programs and indicated that they had sufficiently internalized the robotics concepts to create meaningful robotic avatars of themselves. The children were able to narrate and explain the programs they had created to make their robots stop and perform actions that represented experiences they remembered from the school year. For example, one girl explained that "In September, on the first day of gym, [my robot] shakes…It's not letting anyone tag it." Through her narrative, this child demonstrated her understanding that her robot's programmed actions represent a memorable event in one of her favorite classes, and that this event can be narrated like a story.

Another student demonstrated her conceptual integration of the robotic program and her self-representation by personally acting out some of her robot's actions as she watched it move, such as shaking her head with the robot's shaking motion. Her creative narration also indicated a thoughtful intentionality in the program she created ("So in September, I was looking for my name [tag] so I turned. I put my light on and off because I was like, 'Bingo'"). Her use of the first person for her robotic avatar emphasized her internalization of its representational function.

The programs represented a creative combination of actions and emotions (Libman, 2011). In addition to the concrete use of actions, such as "sing" to represent singing a favorite Hebrew song, one student used "spin" to show her excitement at meeting a favorite teacher, while another student's robot moved forward and backward to represent her rolling pin flattening dough for matzah. A few exceptional programs were even more abstract, representing students' emotions using symbolism and imagery. For example, one boy's robot shook and then spun to represent the overwhelming experience of tasting a bitter lemon during Citrus Fruits Explorations. Another child created a program to turn the robot's lights on and off at the stop for her birthday celebration to represent the metaphoric idea of her eyes "lighting up" in excitement.

Overall, one of the most challenging aspects of the project emerged after all three stops were programmed separately: the children had to compile their three programs into one long journey, arrange their events chronologically and get their robots to travel the correct distance (and in a straight line) between each event so that it stopped at exactly the right months. This required extensive trial-and-error, with the children programming a certain number of forwards, testing out their robot with the line of posters representing the months of the year, and modifying the programs accordingly.

This task proved quite challenging for many of the Kindergarteners, because they did not yet possess the estimation skills to guess how many forwards would correspond to a certain distance for the robot to travel. An unrelated complication was that the robots themselves did not travel exactly in a straight line, and required the children to closely follow them, nudging them in the right direction every few seconds. Between hitting the wall or being nudged too forcefully, robot wheels and platforms would often fall off in the testing phase, which caused significant delays in the program revision process.

On the final presentation day, parents and relatives of each child were given a handout of the child's explanations of each event and program. The handout considerably helped to orient family members, as many of the Kindergarteners were not yet able to coherently verbalize their projects. The parents were amazed to see their child's robot travel the poster time line "by itself" and the students showed a clear sense of pride and confidence in demonstrating and explaining their robot's behaviors to their families. Interestingly, some students narrated their robot's journey in the third person ("My robot is doing…") while others used the first person ("Now I am going…"). This difference suggested a developmental divide between those Kindergarteners who had grasped the representational concept of the Mi Ani project and fully identified with their robot, as opposed to those who considered their robot a third party that was simply duplicating their past actions.

Post-Curriculum Assessment

Children were assessed on the thoroughness of their understanding and application of core concepts and skills using the TangibleK assessment form, a 6-point Likert scale. See Table 2 for the interpretation of each point of the scale.

This assessment tool evaluates concepts and skills derived from programming the robots as part of the TangibleK curriculum. For each of the curriculum's six activities (defined in Table 1), the children were presented with a series of tasks involving either verbal explanation or physical manipulation of the robots or the programming blocks. Their responses were systematically coded using the above Likert scale for demonstrated understanding and implementation of a set of core concepts relevant to each particular curriculum unit. For example, for the "Loops and

Table 2. Definition of the six points of the Likert scale used in the TangibleK assessment form

5	4	3	2	1	0
Complete Achievement of goal/task/understanding	Mostly Complete Achievement of goal/task/understanding	Partially Complete Achievement of goal/task/understanding	Very Incomplete Achievement of goal/task/understanding	Did Not Complete goal/task/understanding	Did not attempt/Other

Parameters" unit, a student might be asked to arrange a set of programming icons to tell the robot to turn four times. To code the student's response, a Likert score would be determined for each of several core concepts, including "Knows when and how to use Repeats," "Selects the right instructions," and "Arranges instructions in the correct order." If applicable, the student's debugging skills for each set of tasks would be evaluated by generating Likert scores for core concepts such as "Recognizes incorrect instructions" and "Attempts to solve the problem."

Many children in the class achieved a high level of comprehension in the application of the powerful ideas related to robotics and programming over the course of the curriculum's six activities and culminating in the Mi Ani project. Regardless of how completely they understood the more abstract concepts of the project, the Kindergarteners showed confidence in their programming skills during post-assessments ("This is so easy"; "I already know I did it right"), even if they actually made a few mistakes in terms of programming syntax. Nonetheless, it was clear that by the completion of the curriculum all of the children understood the logic of the program blocks they were taught.

In the first three curriculum activities, which introduced the engineering design process, robotics, and control flow by sequencing instructions, children's levels of achievement on most skills were particularly high (77% on average achieving "complete" or "mostly complete" understandings). One notable exception was the connection of robotic parts (i.e. connecting tiny wires) which was more difficult for small hands (achievement level was 44% on this task). In the last three activities, which introduced the more sophisticated programming concepts of loops and parameters, sensors, and branches, fewer children (41%, on average) attained the same high level of understanding.

For instance, more children achieved the highest scores on properly sequencing instructions when the programming activities involved only action instructions (71%) as compared to activities which necessarily involved the conceptually and functionally more complicated control flow instructions, or sensors (44%). Programs that use special control flow instructions look linear, but the robot does not carry out one action per programming block, as it does with a program containing only action instructions. This introduces a conceptual complexity to programming with control flow instructions that does not exist with action instructions alone. These differences in achievement reflect the relative theoretical difficulty of each programming structure, with conceptually simpler structures being more conducive to understanding than more complex ones.

The fact that fewer children achieved the highest levels of understanding on more complex topics than on the introductory concepts might indicate that more time is necessary for children to explore the harder material in order to fully understand it. Exposure to concepts makes a difference. For example, after initial introduction to "repeat" instructions, 38% of children achieved a high level of understanding, compared to after their final Mi Ani projects where 86% of children achieved a high level of understanding. Scores on correspondence (matching a programming instruction to intended robot actions) increased from 60% at the curriculum's introduction to 94% after completing the Mi Ani project. Scores on both correspondence and sequencing decreased as activities introduced increasingly complex programming concepts – correspondence decreased to 36% and sequencing decreased to 39%. However, sequencing scores in the Mi Ani project saw a similar jump to that of correspondence scores, with 100% of children achieving a high level of understanding. During the open ended Mi Ani projects, children were able to choose a level of challenge that matched their abilities, leading to a program that was comfortable for each child and led to maximum success.

Analyzing the Experience through the PTD Theoretical Framework

The use of technology in the classroom should be guided by a particular pedagogical stance or theoretical framework. In the Mi Ani project, the PTD (Positive Technological Development) framework guided the design of the educational experience. PTD is a natural extension of the computer literacy and the technological fluency movements that have influenced the world of educational technology in the last thirty years by adding psychosocial and ethical components to the cognitive ones (Bers, 2008; Bers, 2006; Bers, 2010a). PTD provides a model for developing and evaluating technology-rich programs.

From a theoretical perspective, PTD is an interdisciplinary approach that integrates ideas from the fields of computer-mediated communication, computer-supported collaborative learning, and constructionist learning and views them in light of research in applied development science and positive youth development. Informed by both Constructionism (Papert, 1993) and Positive Youth Development (Phelps et al., 2009), PTD is a multi-dimensional framework that proposes six C's of positive behaviors supported by new technologies: content creation, creativity, communication, collaboration, community building and choices of conduct (Bers, 2010b). As a framework to guide the design and implementation of educational interventions, PTD takes into consideration the learning environment and the pedagogical practices, as well as cultural values and rituals, that mediate teaching and learning (Rogoff, 2003; Rogoff, Turkanis & Bartlett, 2001). The following paragraphs summarize the Mi Ani experience in light of the 6 C's of PTD.

Content creation is strongly supported by the CHERP software, with which children were able to program robots through either a graphical or tangible user interface. With directed challenges in the early phase of the curriculum and then in dealing with their own goals for their Mi Ani projects, the children developed competence in the technological domain and learned valuable problem-solving skills, logical thinking, and how to debug and revise their programs, while engaging in developing computational thinking.

The open-ended nature of the Mi Ani project was successful in fostering *creativity* in the ways the children used the technology. Although explicit lessons and challenges were initially helpful in teaching and demonstrating uses of robotics, they offered limited outlets for creativity in the solutions. However, children found extremely creative ways to express their identities and experiences when given the freedom to do so. As they had time to creatively explore and gain competence in navigating CHERP, the Kindergarteners also gained a strong sense of confidence in their programming abilities, which was extremely evident in their pride as they demonstrated their projects to their families. The students' growing confidence also bolstered their ability to overcome technological frustrations when their robots malfunctioned or when CHERP froze, either by trying to fix the problem themselves or soliciting help.

The TangibleK curriculum did support *communication* in several ways. First, the culminating Mi Ani project communicated to parents and family a sense of identity. Kindergarteners programmed their robots as personal avatars to communicate their views of themselves as members of the Jewish people and their particular classroom. Second, children communicated with each other to overcome frustrations and share possible solutions. The Kindergarteners also engaged in *collaboration* by working with their peer partners to solve the robotics challenges posed to them at the beginning of the curriculum. With a wide spectrum of developmental levels and social skills children not only learned about robotics but also about team work. The final projects were shared in an open house. This served a *community building* purpose.

Finally, many aspects of the robotics curriculum at JCDS allowed the Kindergarteners to make *choices of conduct*, which offer opportunities to build character. During the creation of their platforms for the Mi Ani project, children were encouraged to examine and reflect upon their defining character traits. In addition, the limited supply of materials created social challenges in which the students had the freedom to choose whether or not to act responsibly and share materials. During the first phase of the curriculum, children had to navigate sharing of responsibilities and resources with their partners, and negotiate turn-taking. In the rug area, children at the Lego bins could choose to take an excessive amount of Lego, or to take only what they needed so they could share the Lego with their peers.

CONCLUSION

This chapter describes an innovative approach, the TangibleK robotics program, to bring ideas of computer science and engineering into the early childhood classroom in a developmentally appropriate way. Furthermore, the experience described in this chapter focuses on how robotics can be used as a tool for identity exploration. By creating robotic representations of themselves, young children underwent a process of examining their beliefs and practices as members of the Jewish community in a pluralistic school. The dynamic nature of the technology allowed for expressions of actions and experiences, as opposed to only static symbols or facts.

We believe that robotics could be beneficially integrated into any receptive Kindergarten curriculum, if accomplished in a manner that is sensitive to the needs and abilities of young children. There exists some controversy as to how suitable such programs are for this grade level, and we stress that the appropriate design of the robotics technology and the integration with other areas of the curriculum is crucial to its success in the classroom. According to our data, the majority of the *Gan Nitzan* Kindergarteners did achieve high understanding on average for basic programming skills (control flow by sequencing intructions), although most had not fully grasped the more sophisticated skills of loops and parameters, sensors, and branches by the end of the curriculum. The data suggests that greater time exposure to the more complex concepts may have improved the Kindergarteners' level of understanding in these areas.

An ethnographic approach proved particularly useful in holistically evaluating the TangibleK curriculum in this context. While quantitative data demonstrates the effects of the robotics technology within the cognitive domain, the ethnographic study elucidated the potential value of the technology across less quantifiable domains such as creativity and identity exploration. Our ethnographic perspective also acknowledges the role of culture and student population on the implementation of the project. We anticipate the importance of shaping future robotics curricula around the unique culture and population group in each classroom. We recognize that the current project was carried out within a relatively homogenous group of predominantly white, middle-class, Jewish children from the greater Boston area. It is likely that these children had received significant exposure to computer technology at home and it was observed that they were all at least beginning readers by the time the program began. We imagine that these are significant factors that contribute to successful navigation of the graphical TangibleK interface and understanding of syntax and sequential nature of the programs. Such technological fluency and literacy cannot be expected of all populations of kindergarteners, such as those coming from families of lower socio-economic status, and this must be taken into account when designing the curricula. One focus of future research is the development of more affordable robotic hardware that is developmentally appropriate and that can be used with the CHERP programming environment.

In our multicultural world, projects such as this one, that offer children opportunities to explore and represent their own dynamic notion of identity, present educators and researchers with a lens into young children's conceptualization of their identity. While the experience described in this chapter was carried out with a specifically Jewish population, it is our hope that this project will be replicated with a wide range of cultural, religious, or ethnic groups.

ACKNOWLEDGMENT

The authors would like to extend many thanks to the wonderful and creative teachers who used our curriculum in their classroom for this study: Jared Matas and Miriam Newman; to their principal Ruth Gass for embracing a new venture into JCDS; to Nehama Libman for her thesis work on the Mi Ani project, and to all the students in the DevTech group, specially Louise Flannery, Elizabeth Kazakoff and Jordan Crouser for making this project possible. This research was funded by the National Science Foundation (NSF Grant DRL-0735657). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Bers, M., & Horn, M. (2010). Tangible programming in early childhood: Revisiting developmental assumptions through new technologies. In Berson, I. R., & Berson, M. J. (Eds.), *High-tech tots: Childhood in a digital world* (pp. 49–70). Charlotte, NC: Information Age.

Bers, M., Rogers, C., Beals, L., Portsmore, M., Staszowski, K., & Barnett, M. (2006). *Early childhood robotics for learning*. Poster Session at the International Conference on the Learning Sciences. Bloomington, IN.

Bers, M. U. (2006). The role of new technologies to foster positive youth development. *Applied Developmental Science*, *10*(4), 200–219. doi:10.1207/s1532480xads1004_4

Bers, M. U. (2008). *Blocks to robots: Learning with technology in the early childhood*. New York, NY: Teachers College Press.

Bers, M. U. (2010a). When robots tell a story about culture...and children tell a story about learning. In Yelland, N. (Ed.), *Contemporary perspective on early childhood education* (pp. 227–247). Maidenhead, UK: Open University Press.

Bers, M. U. (2010b). The TangibleK robotics program: Applied computational thinking for young children. *Early Childhood Research and Practice, 12*(2).

Bers, M. U., & Urrea, C. (2000). Technological prayers: Parents and children exploring robotics and values. In Druin, A., & Hendler, J. (Eds.), *Robotics for kids: Exploring new technologies for learning* (pp. 194–217). New York, NY: Morgan Kaufman.

Blikstein, P., Buechley, L., Horn, M., & Raffle, H. (2010). A new age in tangible computational interfaces for learning. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences,* (vol. 2, pp. 130-132).

Bodrova, E., & Leong, D. (2007). *Tools of the mind: The Vygotskian approach to early childhood education* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Cejka, E., Rogers, C., & Portsmore, M. (2006). Kindergarten robotics: Using robotics to motivate math, science, and engineering literacy in elementary school. *International Journal of Engineering Education*, *22*(4), 711–722.

Horn, M. S., Crouser, R. J., & Bers, M. U. (2011. (in press). Tangible interaction and learning: The case for a hybrid approach. [special issue on Tangibles and Children]. *Personal and Ubiquitous Computing*.

Kazakoff, E., & Bers, M. (2010). *Computer programming in kindergarten: The role of sequencing.* Presentation at ICERI 2010 International Conference of Education, Research and Innovation, 15-17 November 2010, Spain: Madrid.

Libman, N. (2011). *Mi ani? (Who am I?): Robotics as a medium to express Jewish identity.* Master's thesis, Tufts University, Medford, MA.

Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York, NY: Basic Books.

Papert, S. (1993). *The children's machine: Rethinking school in the age of the computer*. New York, NY: Basic Books.

Petroski, H. (2003). Early education. *American Scientist*, *91*, 206–209.

Phelps, E., Zimmerman, S., Warren, A. A., Jelicic, H., von Eye, A., & Lerner, R. M. (2009). The structure and developmental course of positive youth development (PYD) in early adolescence: Implications for theory and practice. *Journal of Applied Developmental Psychology*, *30*(5), 571–584. doi:10.1016/j.appdev.2009.06.003

Repenning, A., Webb, D., & Ioannidou, A. (2010). Scalable game design and the development of a checklist for getting computational thinking into public schools. *SIGCSE 2010: Proceedings of the 41st ACM Technical Symposium on Computer Science Education.*

Resnick, M. (2007). Sowing the seeds for a more creative society. *Learning and Leading with Technology*, (December/January): 18–22.

Resnick, M., Maloney, J., Monroy-Hernandez, A., Rusk, N., Eastmond, E., & Brennan, K. (2009). Scratch: Programming for all. *Communications of the ACM*, *52*(11), 60–67. doi:10.1145/1592761.1592779

Rogers, C. B., Wendell, K., & Foster, J. (2010). The academic bookshelf: A review of the NAE report, "engineering in K-12 education. *Journal of Engineering Education*, *99*(2), 179–181.

Rogoff, B. (2003). *The cultural nature of human development*. New York, NY: Oxford University Press.

Rogoff, B., Turkanis, C. G., & Bartlett, L. (2001). *Learning together: Children and adults in a school community*. New York, NY: Oxford University Press.

Wing, J. (2006). Computational thinking. *Communications of the ACM*, *24*(6), 33–35. doi:10.1145/1118178.1118215

KEY TERMS AND DEFINITIONS

CHERP: Creative Hybrid Environment for Robotic Programming – A hybrid tangible programming system composed of interlocking wooden blocks specifically designed to meet the developmental needs of young children.

Ethnographic: The scientific description of the customs of individual peoples as observed from within the culture.

Graphical Interface: A representation displayed on a computer screen for communicating instructions and feedback.

Identity: The salient characteristics defining an individual's sense of self.

Positive Technological Development (PTD): An interdisciplinary educational approach that integrates ideas from the fields of computer-mediated communication, computer-supported collaborative learning, and constructionist learning and views them in light of research in applied development science and positive youth development.

Tangible Interface: A physical modality for communicating instructions and feedback.

TangibleK: A curriculum created by the Tufts University DevTech Team using developmentally appropriate methodology to teach robotics and computer programming concepts to kindergarten students.

This work was previously published in Robots in K-12 Education: A New Technology for Learning, edited by Bradley S. Barker, Gwen Nugent, Neal Grandgenett and Viacheslav I. Adamchuk, pp. 168-184, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 106

Prototyping of Robotic Systems in Surgical Procedures and Automated Manufacturing Processes

Zheng (Jeremy) Li
University of Bridgeport, USA

ABSTRACT

The prototyping and implementation of robotic system is a scientific and technological integrating of robotic system design, development, testing, and application. This chapter describes the recent development and applications of robotic systems to surgery procedures in biomedical engineering and automated manufacturing processes in industry. It includes the design and development, computer-aided modeling and simulation, prototype analysis, and testing of robotic systems in these two different applications.

1. INTRODUCTION

Robotic system is constructed mechanically of kinematic chains and actuators which function as muscles that transfer potential energy to body motion (Dylan, 2005, p.797 & Schreuder, 2010, p.253). The robotic mechanisms can be controlled to perform different tasks. The applications of programmable logic control to the industry bring revolution for the manufacturing techniques. It allows more sophisticated, flexible, reliable, and cost-effective manufacturing process controls (Feil-Seifer, 2007, p.425 & Kim, 2007, p.13 & Yang, 2008, p.79). Robotic system is to use control system to reduce human labor intervention during manufacturing processes and productions. It plays very important role and puts strong impact in today's industries (McComb, 2003, p.435 & Pott, 2005, p.101). Computer-aided engineering design can quickly model the robotic systems and speed the design and development cycles (Kumar, 2011, p.994 & Menzel, 2000, p.180 & Tsagarakis,

DOI: 10.4018/978-1-4666-1945-6.ch106

2003, p.21). Computer aided manufacturing can improve the engineering integral processes of robotic design, development, engineering analysis, and production (Schiff, 2005, p.122 & Tapus, 2007, p.15 & Tholey, 2003, p.158). The current economic globalization requires significant labor cost reduction through industrial automation, applied robotic system, improved machine tools, and efficient production process (Ahmed, 2009, p.431 & Siebert, 2004, p.127). This chapter aims at the introduction of some robotic systems applied in surgical procedure and automated manufacturing processes. Two new research projects of robotic systems are introduced in this chapter through design analysis, computer modeling, computational engineering simulation, and prototype testing.

2. BACKGROUND

Robotic surgery has many advantages including minimally invasive surgical process, narrow incisions, decreased infection, reduced pains, and less hospital stays (Cadierre, 2001, p.1467 & Eirik, 2009, p.77 & Taylor, 2003). The development of surgical robots can improve the situation limited by current laparoscopic surgery and technologies (Backes, 2008, p.97 & Camarillo, 2004, p.188 & Hu, 2002 & Kazanzides, 2008). Also the robotic surgical system can be set up whereby the patients could be loaded into a vehicle by robotic surgical equipment and surgery can be performed by a surgeon remotely at a nearby mobile advanced surgical hospital (Brown, 2007, p.253 & Estey, 2009, p.488 & Ghomi, 2010, p.87 & Mataric, 2007, p.1). The robotic arms can be manipulated through surgeon's voice instructions to control the endoscopic cameras (Bargar, 2007, p.31 & Carigan, 2007, p.179 & Gerhardus, 2003, p.242 & Tapus, 2008, p.169). Figure 1 shows one robotic surgical application in gallbladder surgery. The gallbladder removal by robotic surgery is minimally invasive by way of robot technology permitting optimal viewing of the surgical field through small incisions with less pain and faster recovery time for patients (Gockley, 2006, p.150 & Hanna, 2011, p.761 & Harja, 2007, p.365 & Rosen, 2011). It closely mimics the surgery that is used in traditional 'open' procedures, but allows surgeon to perform the operation by da Vinci™ Surgical Systems through small incisions that are associated with minimally invasive and laparoscopic surgeries (Gortchev, 2010, p.153 & Kaouk, 2009, p.181 & Koh, 2011, p.1945 & Puntambekar, 2009, p.259).

Figure 2 shows the da Vinci robotics surgical system for gynecology, urogynecology, urology and cardiology procedures. It has flexible, safe, précised features that allows surgeon to operate the surgery with reduced trauma to the patients and a faster recovery time (Kwartowitz, 2006, p.157 & Li, 2002, p.90 & Melvin, 2003, p.33 & Passerotti, 2006, p.193 & Peters, 2007, p.179).

Figure 3 displays another robotic-assisted surgical system to help surgeon in surgery. It allows the surgeon to perform surgery with enhanced precision that reduces post-operative complications, less pain and discomfort after surgery, less scarring and a shorter hospital stay (Kypson, 2004, p.87 & Melvin, 2003, p.11 & Passerotti, 2006, p.193 & Patel, 2011, p.423).

Figure 1. Robotic surgery for gallbladder removal (Gockley, 2006)

Figure 2. da Vinci robotics surgical system (Passerotti, 2006)

The common robotic systems in manufacturing processes include robotics systems for picking and placing work pieces, assembling operations, and handling machine tools (DeLaurentis, 2002, p.91). The common type of robotic unit is the mechanical arms being used in industry and manufacturing processes (Desforges, 2011, p.963). This mechanism, like human arm, can perform many different motions, including left-and-right, up-and-down, and angular motions. Robotic arms can be either controlled through computer program or human interference. Industrial robots have a multi-linked manipulating arm and end effecter such as a gripper mechanism. It can be applied to industrial packaging, welding, painting, printed circuit boards, hazard environment, and many other manufacturing processes (Torbjorn, 2009, p.87). Figure 4 shows one robotic system for picking and placing parts in automobile production. It can increase the efficiency at multiple working cells by automatically picking and placing product components in production (Ullman, 2002, p.178).

Figure 3. da Robotic-assisted surgical system (Melvin, 2003)

Figure 5 shows one heavy duty robot applied in product lifting operation. It has a payload capacity of 2200 lbs and a reach of 10 feet for extremely heavy lifting and manipulation applications in industries of automotive, foundry and building materials (Yang, 2011, p.1010).

Figure 6 displays one robotic system in material removal operations. It assists to process the material removal processes including drilling, milling, polishing, grinding, buffing and deflashing (Yun, 2011, p.977). The robot-assisted manufacturing can boost manufacturing efficiency throughout industrial applications.

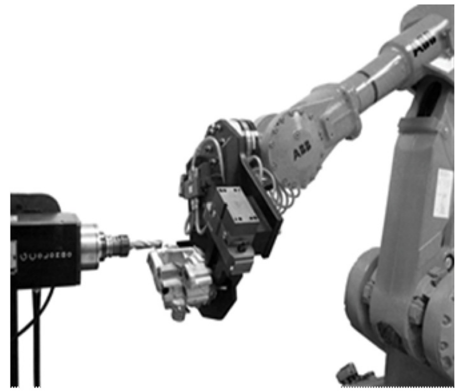
Figure 4. Robotic system for picking and placing parts in automobile production (Torbjorn, 2009)

Figure 5. One heavy duty robot in product lifting and manipulation operation (Yang, 2011)

3. MECHANISM AND CONTROL OF ROBOTIC SYSTEMS

The robotic systems convert the energy into motion. It may include electric motors, linear actuators, piezo motors, air wires, and electroactive polymers (Howard, 2006, p.1360). Electric motor is the most popular type of actuator (Shakernia, 2002, p.2793). Linear actuator is used in applications where the large forces are required in linear motion (Dagalakis, 2007, p.180, Mellinger, 2010). Piezo motor uses the vibration of the piezo elements to drive the motor. Air wire is a special tube and its length can be adjusted as internal air is exerted (Matsui, 2005, p.47). Electroactive polymer, functions as the arm of robotic system, is a special plastic material that can change its length significantly by electricity. Current robotic systems receive tactile signals with sensors simulating the mechanical properties and touch receptors of human fingertips (Gurdan, 2007, p.288). The sensor can consist of a rigid core with conductive fluid surrounded (Kazerooni, 2005, p.3459 & Kennedy, 2003, p.2106). The conductive fluid is enclosed by the elastomeric materials. Electrode is installed at the surface of the rigid core and connected to an impedance-measuring device inside the core.

Figure 6. Robotic system for material removal processes (Yun, 2011)



When the elastomeric material touches an object, the fluid volume surrounded electrode is changed, causing the impedance changes that simulate the loads received from the object. Computer-aided vision is the technology of mechanism that gets data information from images with different format including video sequences or cameras (Lerman, 2006, p.225). It uses the image sensor to detect electromagnetic radiation of visible or infra-red light.

Robots can manipulate objects by “end effectors” (hands) and “manipulators” (arms) (Eljamel, 2007, p.372 & Frutiger, 2002, p.619 & Hong, 2006, p.81). Some robotic arms equipped in exchangeable effectors with which it can perform different tasks. Mechanical grippers are the most common effectors with two fingers that can open and close to pick up and place the objects. Vacuum grippers are simple and common effectors to pick and place large objects through suction. Robotic system control process is complex and dynamic related (Frisoli, 2005, p.195). Some technical methodologies include: (1). zero moment point technology—it is the algorithm used by robotic systems to keep the total forces balanced with combination of gravity, acceleration and deceleration of object body; (2). Hopping—it is the technology that keeps robotic system functioning properly and the algorithm in the control program is generated to maintain the system balance if robotic system is unbalanced; (3). dynamic balance—it is a more advanced methodology to control robotic system by dynamic balancing algorithm and more robust than zero moment point technology; (4). passive dynamics—it is applied with high efficiency in momentum of robotic system and the robotic system only requires less motor power if compared with other conventional robotic systems.

The control of robotic systems consists of three different phases: (1). Perception—sensors are applied to detect the external signal information for actuators; (2). Processing—it translates sensor information into commands to drive actuators; (3). Action—object pattern verification and computer vision can be used to track objects (Gerkey, 2004, p.939). The study of motion in robotic systems includes both kinematics and dynamics. The position, orientation, velocity, and acceleration of end effectors can be calculated through direct kinematics. The kinematics can be used to analyze the same robotic motion in different possibilities to avoid the collision and singularity in robotic systems. With all results from kinematics analysis, dynamics can be applied to study the effect of forces on motion of robotic systems. The dynamic technique is a powerful tool in computer-aided modeling and simulation of robotic systems to improve the control algorithms of a robotic system (Tinelli, 2011, p.2654).

4. NEW RESEARCHES IN SURGICAL AND INDUSTRIAL ROBOTIC SYSTEMS

Based on author's current researches, two new robotic systems have been introduced and analyzed in this chapter including robotic system design and development in surgical procedures and automated manufacturing processes. These two robotic systems are designed in such way that they can be made cheaper and simple without losing the reliability. These robotic systems have been designed and analyzed by computer-assisted modeling, simulation and analysis. The simple prototypes were being built and tested to verify its performance. The preliminary results indicated that these robotic systems function properly during the prototype testing.

4.1 New Robotic System for Surgical Procedures

The surgical operations can be performed through robotic systems with more accuracy. Robotic surgery applies the surgical robots that lead direct or indirect involvement of surgeon in surgical operation. The robotic systems are computerized mechanical systems with motorized mechanisms. The basic robotic systems may consist of sensors, programmable systems, and actuators (Carigan, 2007, p.185). The important advantages of robotic surgery over conventional surgery include the accuracy of surgery, capability of repeating identical motions, and options of remote surgery (Finkelstein, 2010, p. 42). There are three categories in robotic surgery based on different surgeon involvement (Hanna, 2011, p.780): (1). supervisory—before surgical procedure, surgeon inputs the computer program into robotic system so that the surgical operation can be solely performed without surgeon direct interaction, but It requires complex program for the surgery and more advanced technologies to get images and data of patient body; (2). Telesurgery—it is a remote surgery in which surgeon directly manipulates the robotic systems during the surgical process without pre-programming the robotic system and surgeon can perform surgical operation from other locations by using real-time data information obtained from sensors in robotic system; (3). share-control—this procedure needs the most involvement of surgeon who performs the surgical operation by using robotic system that provides more accurate and stable manipulations of surgical instruments. A new surgical robotic system, shown in Figure 7, is designed and de-

Figure 7. New robotic system for surgical application

veloped based on computer modeling, analysis and prototype testing.

Computer tomography scans can get cross-section images that are important to the diagnosis and surgeons can use these images to decide surgical pathways to precisely guide operation instruments with less surgical injury during procedures. The robotic arms in this robotic system can precisely access the patient's body through tiny and narrow incisions. Surgeons can use remote manipulation device to perform complex surgical operation with accuracy and less fatigue because of proper geometrical design, wide angular motion, and 3-dimensional clear visualization. The wide view of surgical operations using this surgical robot provides surgeons with great capability to perform many surgical tasks including robotic prostate cancer surgeries for cancel removal, damage control, sexual and urinary function recovery with minimal side effects. This robotic surgery can benefit patients by quick recovery, less postoperative pain, short hospital stays, small tissue trauma, and better cosmetics. This surgical robot provides surgeons with more precise, flexible and accurate procedures than traditional surgical techniques. It also permits highly efficient suturing because of accurate placement and well controlled motion of surgical instrument in this robotic system. In summary, this surgical robot has the following features:

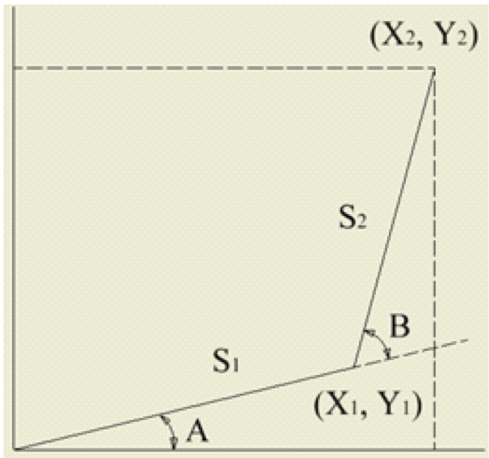
1. The simpler system design allows loose manufacturing tolerance control on major components in this robotic system so that it can be made cheap and maintained easily in the future;
2. The moving components in this robotic system are designed to keep its proper geometry and light weight to benefit surgeon in surgical operation with less moving vibration and higher energy efficiency.
3. Manipulated by surgeons, the robotic arms can have wider angular degrees of motion freedom with more flexible function;
4. Larger visual magnification for better surgical function;
5. Less post-operative length of stay due to the capability of performing minute surgical incisions more precisely.

4.1.1 Mathematics and Computer-Aided Simulation of Robot Arms

The robotic systems are programmed with mathematics that controls the motion of robot arms accurately and reliably. The double-link robot arm is simple, shown in Figure 8, if compared with more complicated triple-link robot arm which involves more difficult mathematics. In the double-link arm system, the first link (assistant arm) turns around the origin of an XY Cartesian coordinate system and the second link (arm to hold the surgical tools) turns around the pivot point which connects link 1 and 2. The angle between link 1 and horizontal axis is defined as A, and angle between link 1 and 2 is assumed as B. The robot arm is at the end positions (X_1, Y_1) of link 1 and (X_2, Y_2) of link 2.

There are three solution methodologies in this robotic system analysis (Carlone, 2010, p.283): (1).kinematics—determine robot arm location (X_1, Y_1) if the angles A and B are given; (2).inverse kinematics—determine angles A and B if robot

Figure 8. Double-link robot arm



arm location (X_1, Y_1) is given; (3).trajectory generation—determine robot arm new location (X_2, Y_2) by adjusting the angle A and B if current location (X_1, Y_1) is given.

4.1.2 Methodology of Kinematics (McComb, 2003)

Based on Figure 8, we can find the end location:

$$X_2 = S_1 \cdot \cos(A) + S_2 \cdot \cos(A+B) \qquad (1)$$

$$Y_2 = S_1 \cdot \sin(A) + S_2 \cdot \sin(A+B) \qquad (2)$$

Here,

$$X_1 = S_1 \cdot \cos(A)$$

$$Y_1 = S_1 \cdot \sin(A)$$

$$X_2 = X_1 = S_2 \cdot \cos(A+B)$$

$$Y_2 = Y_1 + S_2 \cdot \sin(A+B)$$

4.1.3 Methodology of Inverse Kinematics (McComb, 2003)

Referring kinematics Equations (1) and (2), the problem is nonlinear to determine angles A and B if end location (X_2, Y_2) is given. The algebraic methodology can be applied to solve this nonlinear problem and Equations (1) and (2) are combined to get following equations:

$$\cos(A+B) = \cos(A) \cdot \cos(B) - \sin(A) \cdot \sin(B) \qquad (3)$$

$$\sin(A+B) = \cos(A) \cdot \sin(B) + \sin(A) \cdot \cos(B) \qquad (4)$$

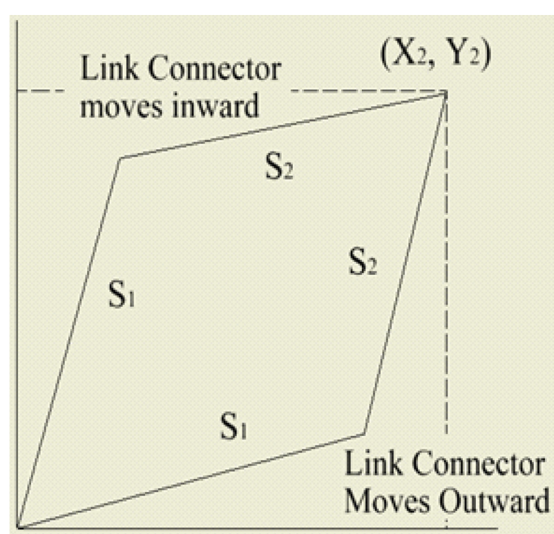
Solving these two equations can derive the new equation of finding angle B:

$$B = \arccos\left(\frac{X_2^2 + Y_2^2 - S_1^2 - S_2^2}{2S_1 \cdot S_2}\right) \qquad (5)$$

Since cos (B) = cos (-B), the inverse kinematics problem has two solutions (link 2 clockwise and counter-clockwise turning), shown in Figure 9.

Angle A can be found based on solution of angle B in inverse kinematics methodology.

Figure 9. Inverse kinematics for robotic arm



$$X_2 = [S_1 + S_2 \cos(B)]\cos(A) - S_2 \cdot \sin(B) \cdot \sin(A) \quad (6)$$

$$Y_2 = S_2 \cdot \sin(B) \cdot \cos(A) + [S_1 + S_2 \cdot \cos(B)] \cdot \sin(A) \quad (7)$$

The angle A can be determined by solving Equations (6) and (7).

4.1.4 Methodology of Trajectory Generation (McComb, 2003)

In case of double-link robot arm, the arm location (X_1, Y_1) is associated to link angles A_1 and B_1. When rotate the arm to a new location (X_2, Y_2), the angles of A_2 and B_2 can be determined through inverse kinematics analysis. There are several ways to control robotic arm moving from old location to a new location: (1). adjust angle A before changing B; (2). adjust angle B before changing A; (3). adjust angles A and B at the same time; (4). robot arm turns clockwise; (5). robot arm turns counterclockwise. A good robot arm design can reduce the energy needed to rotate the robot arm and decrease the time needed to change arm mechanism. Figure 10 shows a robot arm rotating in trajectory generation with angle A changing in a clockwise direction.

Two moving trajectories of robot arm, shown in Figure 10, indicated the desired condition. In straight line trajectory generation methodology, a line or curve can be divided into many small segments and corresponding (X, Y) coordinates at the end of each segment are calculated through computational simulation. For each (X, Y) coordinate pair, the required angles A and B are computed. The robot arm mechanisms are adjusted as each new A and B are calculated. The computer-aided solutions of angles A and B continue, along the line or curve, until the target arm location is reached.

The more complex situation for robotic arm motion could be analyzed through computer-aided modeling and simulation. Figure 11 shows the trajectory curve of robot arm using multiple data points along the moving path. If the angles of A and B change simultaneously, the moving path is much smoother.

Figure 10. Robotic arm rotates clockwise in trajectory generation

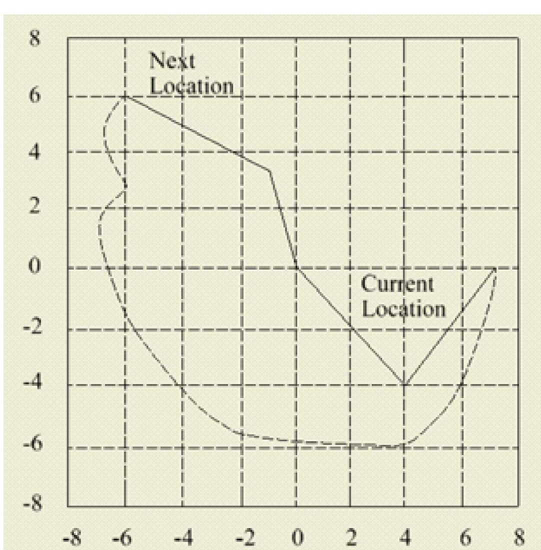


Figure 11. Simulation of robotic arm moving path with multiple data points

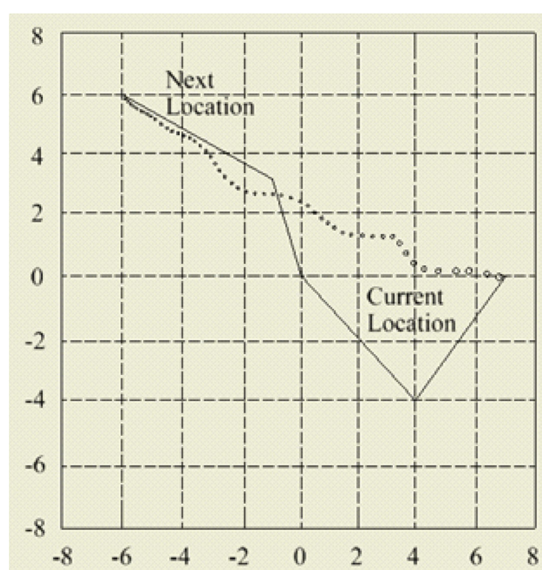


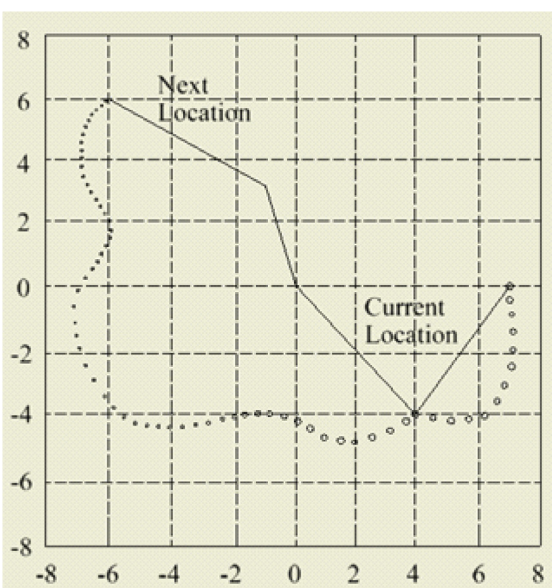
Figure 12 shows the robot arm moving in an alternate path with smoother curve when angles A and B are changed at the same time.

The robot arm system is a nonlinear functioning system which requires the complex mathematical analysis, computer-aided modeling, and computational simulation to analyze and solve the arm motion problems. The variable length links could be used to define the potential coverage regions of robotic arms. Modifying the inverse kinematics methodology can implement the solution for trajectory generation methodology. The computational modeling and simulation are required in the analysis of robotic arm movement.

4.1.5 Prototype Testing

The kinematic motion of this robot is measured by pulse-width modulation (PWM) methodology in preliminary prototype testing. The PWM register's values are used to drive the motors of robotic system in different directions. The curves in Figures 13 (a), (b), (c), (d), (e) and (f) show the power required to drive the movement of this surgical robot arm towards the front, back, left, right, higher and lower directions. All PWM resultant pictures corresponding to the motions of robotic arm in different directions were recorded when robotic system was manipulated by the operator. The average voltage or current values fed to the load is manipulated by controlling the switch connected between the supply and load. The switch quickly changes between on and off status. If switch stays longer in status of "on" than "off", more power will be supplied to the load which can manipulate the robotic system motions. The PWM values in these plots are related to the power that drives the robotic system in different directions.

Figure 12. Simulation of robotic arm moving in an alternate path



The preliminary prototype testing indicated the maximum movement of 16 inches towards the front and back directions, 20 inches towards the left and right directions, and 24 inches towards the higher and lower directions. These preliminary results are in good agreement with design expectation since this surgical robot is expected the motion of 15 × 20 × 24 inches in 3-D design function. More tests have been planned to verify the functionality and reliability of this surgical robot. Further modification on this surgical robot will be made to determine if large moving range can be achieved.

4.2 New Robotic System for Automated and High Speed Viscous Fluid Filling

Figure 14 shows this new robotic system used in automated manufacturing processes. One of the industrial applications is to pick and place the empty cartridges from floor container to the working station in automated viscous liquids filling production line shown in Figure 15.

The precision indexing conveyor controls indexing tolerance to keep part movement within tolerance range. The empty cartridges are picked up and placed to the cartridge holders by this robotic system and securely held in indexing

Figure 13. (a) movement of robot arm towards the front; (b) movement of robot arm towards the back; (c) movement of robot arm towards the left; (d) movement of robot arm towards the right; (e) movement of robot arm towards higher direction; (f) movement of robot arm towards lower direction

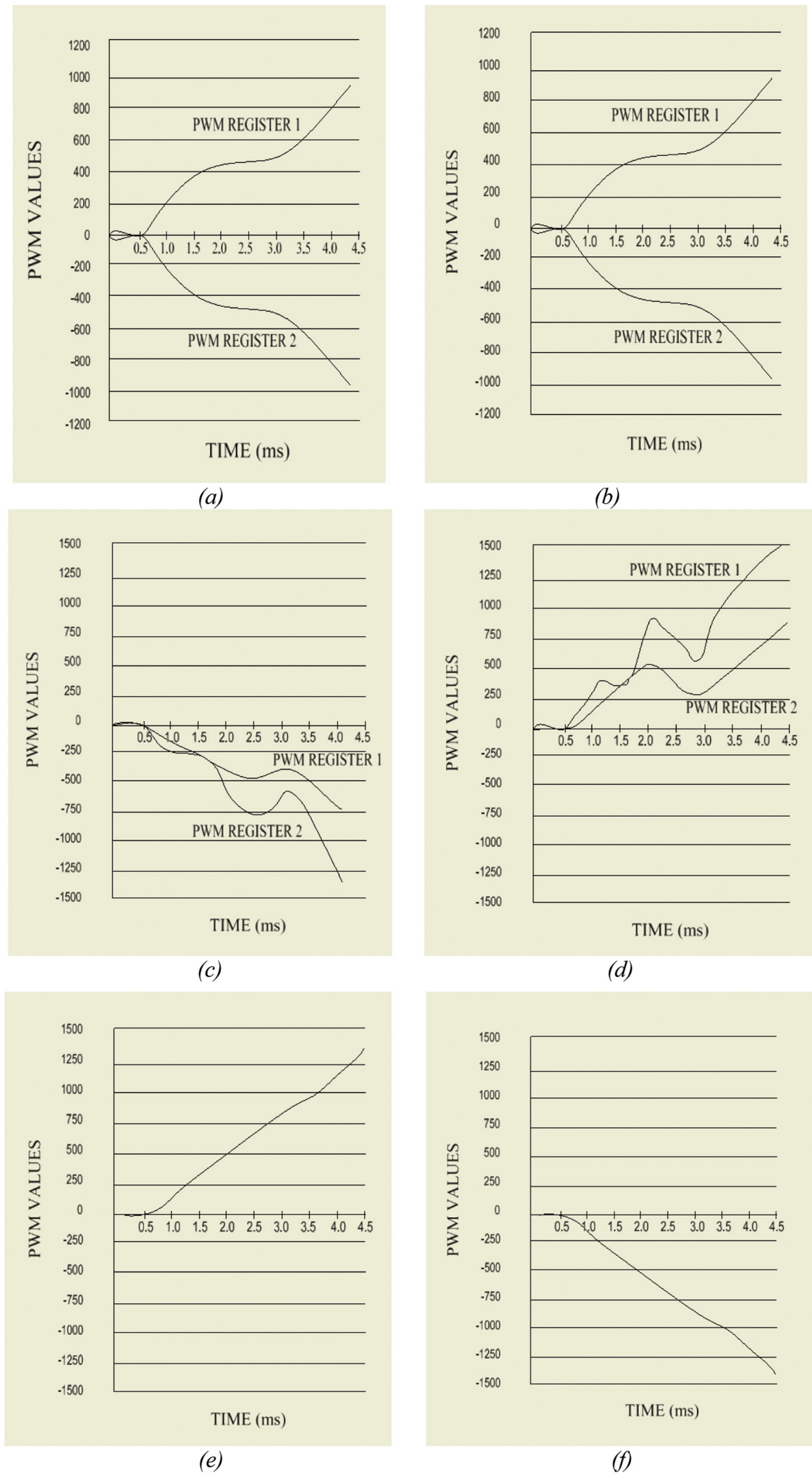
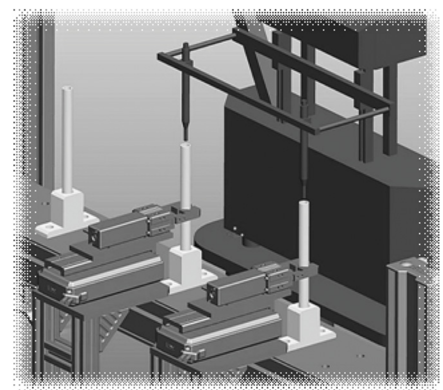


Figure 14. New robotic system for manufacturing applications

Figure 15. Robot-assisted automated and high speed viscous liquids filling production line



conveyor. The liquids filling holders are controlled to keep good dimensional and location tolerances between holders and cartridges to secure the product in automated and high speed filling process. When empty cartridges are delivered to the location below filling nozzles, the indexing conveyors stop through sequential control of PLC system that receives filling holder location signal from opposed mode sensor. Horizontal sliders mounted in gripper system drive gripper pair forward to secure cartridges in filling process. The nozzles are driven downward by vertical sliders and inserted into cartridges to begin filling process. As soon as filling process is completed, the vertical air slide drives nozzles upward. Then grippers release cartridges and move away from filling holder location for next filling cycle. The control system in this prototype can detect each filling pump to precisely determine the amount of liquids being filled. When target filling volume is reached, all pumps and nozzles shut off instantly, resulting in accurate filling process of high viscous liquid products. The control system stores all filling parameters in memory for fast changeovers.

This robot-assisted automated and high speed liquid filling system is accurately and reliably controlled in filling of different high viscous liquids. It can be applied to many industries including pharmaceutical, cosmetic, dairy, chemical, food, etc. For example, some thick liquids, cosmetic creams, and chunky sauces with pasteurized temperatures can all be filled in this new robot-assisted automated filling system. The positive displacement pump will be required in high viscosity and high temperature environment. Double driving rotors are designed in this pump to make this filling system more durable in delivery of high viscosity liquids.

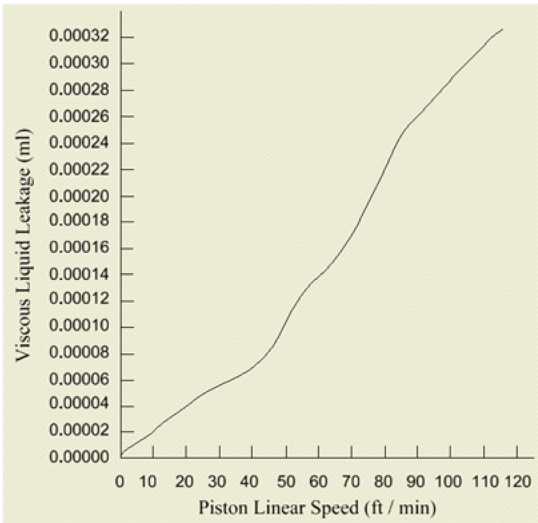
4.2.1 Computer Aided Modeling and Simulation

The computer-aided simulation of liquid leaking rate L during high speed liquid filling process can be determined as follow (Klopcar, 2007, p.86):

$$L = Q \cdot C \cdot \left(\frac{T_b}{P_b}\right) \cdot E^{2.5} \cdot H \cdot \left(\frac{P_1^2 - P_2^2}{I \cdot G \cdot T_a \cdot N_a \cdot J}\right)^{0.5} \quad (8)$$

The computer-aided simulation result, shown in Figure 16, indicated that the liquid leakage in this high speed filling process can be ignored. The computer-aided simulation shows that this new robotic-assisted automated and high speed filling system has reliable function, good sealing capability and high production rates. Also the manufacturing cost for this system is low due to

Figure 16. Viscous liquid leakages vs. piston linear speed



less requirement of tolerance control for assembled robotic components in this new robotic-assisted manufacturing system design.

4.2.2 Prototype Testing

The prototype of this new robotic-assisted automated and high speed liquid filling system has been tested with preliminary results shown in Table 1.

Both prototype testing and computational simulation show the similar results that verify the creditability and feasibility of this new robot-assisted automated and high speed liquid filling system.

Table 1. Liquid leakage vs. piston linear speed

Piston Linear Speed (Ft/in)	Estimated Liquid Leakage (ml)	Piston Linear Speed (Ft/ Min)	Estimated Liquid Leakage (ml)
10	0.000015	70	0.000145
20	0.000035	80	0.000185
30	0.000055	90	0.000205
40	0.000088	100	0.000235
50	0.000102	110	0.000265
60	0.000125	120	0.000305

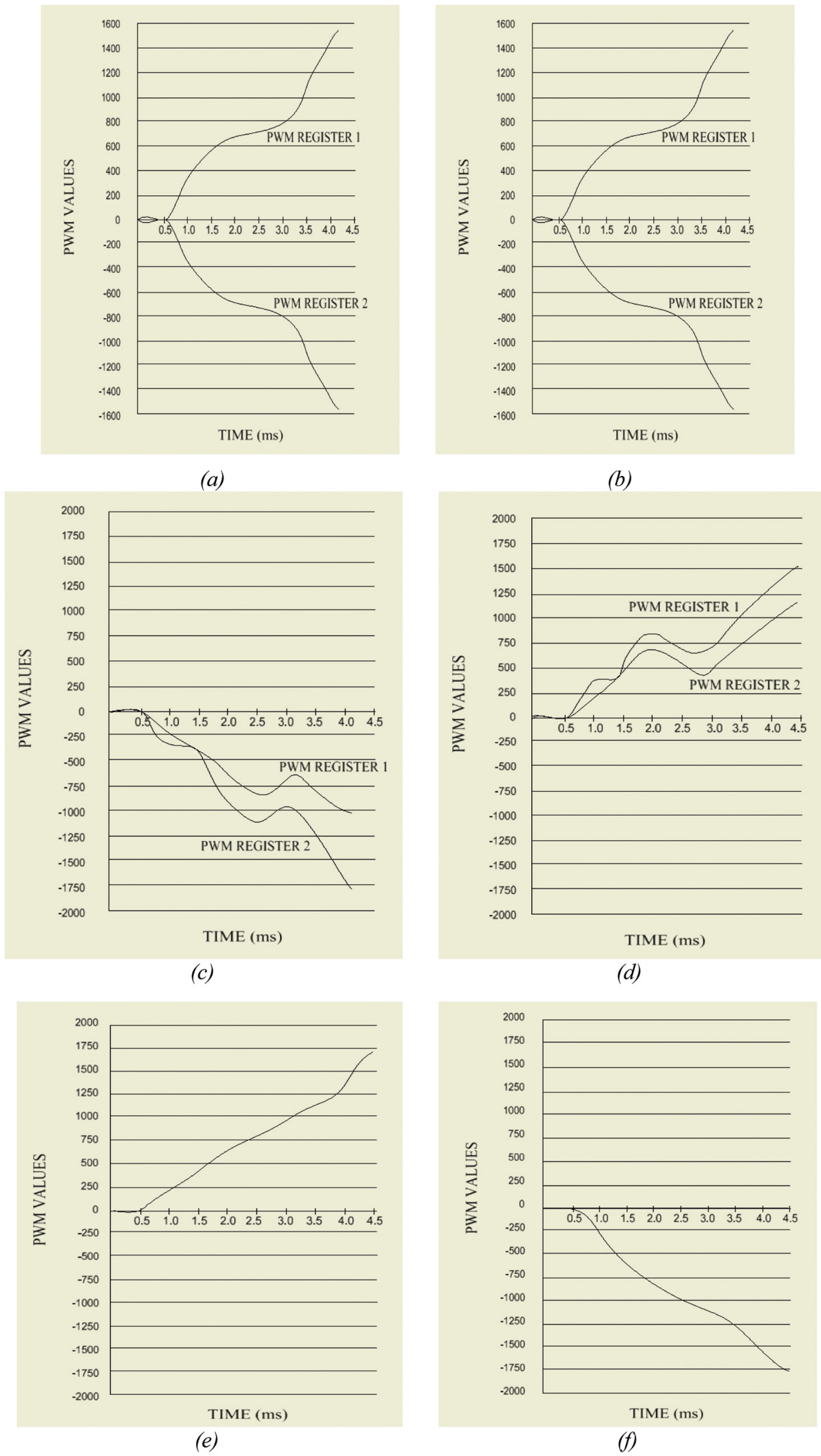
The curves in Figure 17 (a), (b), (c), (d), (e) and (f) represent the power related movement of manufacturing robot arm towards the front, back, left, right, higher and lower directions. The kinematic motion of this robot is measured by pulse-width modulation (PWM) methodology in preliminary prototype testing. The PWM register's values are used to drive the motors of robotic system in different directions. All PWM resultant pictures corresponding to the motions of robotic arm in different directions were recorded when robotic system was manipulated by the operator. The average voltage or current values fed to the load is manipulated by controlling the switch connected between the supply and load. The switch quickly changes between on and off status. If switch stays longer in status of "on" than "off", more power will be supplied to the load which can manipulate the robotic system motions. All these PWM plots represent the power needed to drive the robot arm moving in different directions.

The preliminary testing of this manufacturing robot indicated the maximum movement of 40 inches towards the front and back directions, 48 inches towards the left and right directions, and 52 inches toward the higher and lower directions. The above preliminary results are in good agreement with design expectation because this manufacturing robot is expected to move 3-D dimension in 40 × 49 × 52 inches to meet design target. To verify the functionality and reliability of this surgical robot, more prototype tests will be performed and further modifications of mechanism in this manufacturing robot will be carried out to increase moving range in different directions.

5. FUTURE IMPROVEMENT

The future improvement on these two new robotic systems include: (1). Simplify the robotic systems for easy maintenance and cost-effectiveness, in-

Figure 17. (a) movement of robot arm towards the front; (b) movement of robot arm towards the back; (c) movement of robot arm towards the left; (d) movement of robot arm towards the right; (e) movement of robot arm towards higher direction; (f) movement of robot arm towards lower direction



cluding making arm opening more accessible; (2). Investigate using lighter and durable materials; (3). Modify the robotic mechanism to increase its moving range in different directions; (4). Improve system design for better function including mechanical joint connections to keep more accurate and smooth robotic motion; (5). Optimize the robotic mechanism through computer-aided simulation to verify and guide further design modification; (6). Implement the prototype and prepare to send for field evaluations.

6. CONCLUSION

Robotic system is not only significantly increasing the production speed but also more accurately controlling product quality. It can maintain consistent quality, shorten lead time, simplify material handling, optimize work flow, meet product demand for flexibility and convertibility in production, and perform some specific, precise, and dangerous tasks in industry that are very difficult using human work forces. Robotic surgery has greater levels of accuracy and precision than traditional open surgery. The robotic systems have also been widely used in many industrial fields including automated and high speed production, automobile manufacturing, hazardous duty service, and heavy facility maintenance. Two new research projects of robotic system introduced in this chapter are illustrated mathematically with verifications from computer-aided analysis and preliminary prototype testing. These new research projects can be applied to robotic surgery and automated manufacturing process. Further modifications will keep these robot-assisted systems with more reliable function, flexible features, reduced cost, and enlarged moving range.

REFERENCES

Ahmed, K., Khan, M., Vats, A., Nagpal, K., Priest, O., & Patel, V. (2009). Current status of robotic assisted pelvic surgery and future developments. *International Journal of Surgery*, *7*, 431–440. doi:10.1016/j.ijsu.2009.08.008

Albu-Schaffer, A., Haddadin, S., Ott, C., Stemmer, A., Wimbock, T., & Hirzinger, G. (2007). The DLR lightweight robot: Design and control concepts for robots in human environments. *International Journal of Industrial Robot*, *34*, 376–385. doi:10.1108/01439910710774386

Backes, F., Seamon, L., & Fowler, J. (2008). Robotic radical hysterectomy and pelvic mphadenectomy for uterine rhabdomyosarcoma. *Journal of Robotic Surgery*, *2*, 97–200. doi:10.1007/s11701-008-0095-5

Bargar, W. L. (2007). Robots in orthopaedic surgery - Past, present, and future. *Clinical Orthopaedics and Related Research*, *463*, 31–36.

Brown, W. T., Wu, X., & Wen, B. C. (2007). Early results of CyberKnife imageguided robotic stereotactic radiosurgery for treatment of lung tumors. *Journal of Computer Aided Surgery*, *12*, 253–261.

Cadierre, G. (2001). Feasibility of robotic laparoscopic surgery. *World Journal of Surgery*, *25*, 1467–1477.

Camarillo, D. (2004). Robotic technology in surgery: past, present, and future. *American Journal of Surgery*, *188*, 2–15. doi:10.1016/j.amjsurg.2004.08.025

Carigan, C., Tang, J., Roderick, S., & Naylor, M. (2007). A configuration-sapce approach to controlling a rehabilitation arm exoskeleton. *IEEE 10th International Conference Rehabilitation Robotics,* (pp. 179-187).

Carlone, L., Ng, M., Du, J., Bona, B., & Indri, M. (2010). Simultaneous localization and mapping using rao-blackwellized particle filters in multi robot systems. *Journal of Intelligent & Robotic Systems*, *63*, 283–307. doi:10.1007/s10846-010-9457-0

Chen, P., & Sheng, W. (2011). Transformative CAD based industrial robot program generation. *Journal of Robotics and Computer-Integrated Manufacturing*, *27*, 942–948. doi:10.1016/j.rcim.2011.03.006

Constantin, A. (2006). Evaluating arm movement imitation. *American Journal of Undergraduate Research*, *4*, 9–12.

Dagalakis, N., Kim, Y., Sawyer, D., & Shakarji, C. (2007). Development of tools for measuring the performance of computer assisted orthopaedic hip surgery systems. *Proceeding of the ACM Workshop on Performance Metrics for Intelligent Systems (PerMIS)*, Gaithersburg, (pp. 180–187).

DeLaurentis, K., & Mavroidis, C. (2002). Mechanical design of a shape memory alloy actuated prosthetic hand. *Journal of Health Care*, *10*, 91–106.

Desforges, X., Habbadi, A., & Archimède, B. (2011). Design methodology for smart actuator services for machine tool and machining control and monitoring. *Journal of Robotics and Computer-Integrated Manufacturing*, *27*, 963–976. doi:10.1016/j.rcim.2011.03.004

Dylan, A. S., & Matarić, M. J. (2005). Insights toward robot-assisted evacuation. *International Journal of the Robotics*, *19*, 797–818.

Eirik, L., Johansen, B., Gjelsvik, T., & Lango, T. (2009). Ultrasound based localization of wireless microrobotic endoscopic capsule for the GI tract. *Proceeding of the 21st Conference of the Society for Medical Innovation and Technology (SMIT),* Sinaia, (pp. 77–79).

Eljamel, M. (2007). Validation of a neurosurgical robot using a phantom. *International Journal of Medical Robotics and Computer Assisted Surgery*, *3*, 372–377. doi:10.1002/rcs.153

Estey, E. (2009). Robotic prostatectomy: The new standard of care or a marketing success. *Journal of Canadian Urological Association*, *3*, 488–490.

Feil-Seifer, D., Skinner, K., & Mataric, M. J. (2007). Benchmarks for evaluating socially assistive robotics. *Journal of Interaction Science*, *8*, 423–439.

Finkelstein, J., Eckersberger, E., Sadri, H., Taneja, S. S., & Djavan, B. (2010). Open versus laparoscopic versus robot-assisted laparoscopic prostatectomy: The European and US experience. *Reviews in Urology*, *12*, 35–43.

Frisoli, A., Rocchi, F., Marcheschi, S., Dettori, A., Salsedo, F., & Bergamasco, M. (2005). A new force-feedback arm exoskeleton for haptic interaction in virtual environment. *First IEEE Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems,* (pp. 195-201).

Frutiger, D. R., Bongard, J. C., & Iida, F. (2002). Iterative product engineering: evolutionary robot design. *Proceedings of the Fifth International Conference on Climbing and Walking Robots*, (pp. 619–629).

Gerhardus, D. (2003). Robot-assisted surgery: The future is here. *Journal of Healthcare Management*, *48*, 242–251.

Gerkey, B., & Matarić, M. J. (2004). A formal framework for the study of task allocation in multi-Robot systems. *The International Journal of Robotics Research*, *23*, 939–954. doi:10.1177/0278364904045564

Ghomi, A., & Askari, R. (2010). Use of a bidirectional barbed suture in robot-assisted sacrocolpopexy. *Journal of Robotic Surgery*, *4*, 87–89. doi:10.1007/s11701-010-0188-9

Gockley, R., & Matari'c, M. (2006). Encouraging physical therapy compliance with a hands-off mobile robot. *Proceedings of the 1st Annual Conference on Human-Robot Interaction,* (pp. 150–155).

Goerzen, C., Kong, Z., & Mettler, B. (2010). A survey of motion planning algorithms from the perspective of autonomous UAV guidance. *Journal of Intelligent & Robotic Systems*, *57*, 65–100. doi:10.1007/s10846-009-9383-1

Gortchev, G., Tomov, S., Tantchev, L., Velkova, A., & Radionova, Z. (2010). Da Vinci robotic surgery in the treatment of benign and malignant gynecologic tumors. *Gynecological Surgery*, *7*, 153–157. doi:10.1007/s10397-009-0535-x

Gurdan, D., Stumpf, J., Achtelik, M., Doth, K., Hirzinger, G., & Rus, D. (2007). Energy-efficient autonomous four-rotor flying robot controlled at 1 khz. *Proceeding of the IEEE International Conference on Robotics and Automation.*

Hanna, R., & Boggess, J. (2011). Applications of surgical robotics in gynecologic surgery. *Surgical Robotics*, *4*, 761–789. doi:10.1007/978-1-4419-1126-1_32

Harja, J., Tikkanen, J., Sorvoja, H., & Myllyla, R. (2007). Magnetic resonance aging compatible, three-degrees-of-freedom joystick for surgical robot. *International Journal of Medical Robotics and Computer Assisted Surgery*, *3*, 365–371. doi:10.1002/rcs.159

Hong, S. S., Heo, S. C., Kim, E., & Hang, Y. J. (2006). A study on the practical human robot interface design for the development of shopping support robot. *Journal of Korean Society of Design Science*, *19*, 81–90.

Howard, A., & Gaurav, S. Sukhatme, S., & Matarić, M. J. (2006). Multi-robot mapping using manifold representations. *Proceedings of the IEEE – Special Issue on Multi- Robot Systems, 94,* 1360 -1369.

Hu, T., Castellanos, A., & Tholey, G. (2002). Real-time haptic feedback laparoscopic tool for use in gastro-intestinal surgery. *Fifth International Conference on Medical Image Computing and Computer Assisted Intervention, 12,* (pp. 6-16).

Kaouk, J., & Goel, R. (2009). Robotic-assisted laparoscopic bladder augmentation. *Pediatric Robotic Urology*, *2*, 181–188. doi:10.1007/978-1-60327-422-7_12

Kazanzides, P., Fichtinger, G., Hager, G., Okamura, A., Whitcomb, L., & Taylor, H. (2008). Surgical and interventional robotics: core Concepts, technology, and design. *IEEE Robotics & Automation Magazine*, *15*, 122–130. doi:10.1109/MRA.2008.926390

Kazerooni, H. (2005). Exoskeletons for human power augmentation, *Proceeding of IEEE Conference of Intelligent Robots and Systems,* (pp. 3459–3464).

Kennedy, C. W., & Desai, J. P. (2003). Force feedback using vision. *The 11th International Conference on Advanced Robotics,* (pp. 2106-2111).

Kim, B., Park, J., Kang, B., & Moon, C. (2011). Fabrication and property analysis of a MEMS micro-gripper for robotic micro-manipulation. *Journal of Robotics and Computer-Integrated Manufacturing*, *27*, 1021–1025.

Kim, W. S. (2007). The design of humanoid robot platgorm: HUBO. *Journal of Korean Society of Design Science*, *20*, 13–16.

Klopcar, N., Tomsic, M., & Lenarcic, J. (2007). A kinematic model of the shoulder complex to evaluate the arm-reachable workspace. *Journal of Biomechanics*, *40*, 86–91. doi:10.1016/j.jbiomech.2005.11.010

Koh, C., Tsang, C., & Kim, S. (2011). A new application of the four-arm standard da Vinci® surgical system: Totally robotic-assisted left-sided colon or rectal resection. *Surgical Endoscopy*, *25*, 1945–1952. doi:10.1007/s00464-010-1492-1

Kumar, R., Kalra, P., & Prakash, N. (2011). A virtual RV-M1 robot system. *Journal of Robotics and Computer-Integrated Manufacturing*, *27*, 994–1000. doi:10.1016/j.rcim.2011.05.003

Kwartowitz, D. M., Herrell, S. D., & Galloway, R. L. (2006). Toward image-guided robotic surgery: Determining intrinsic accuracy of the da Vinci robot. *International Journal of Computer Assisted Radiology and Surgery*, *1*, 157–165. doi:10.1007/s11548-006-0047-3

Kypson, P., & Chitwood, W. R. Jr. (2004). Robotic applications in cardiac surgery. *International Journal of Advanced Robotic Systems*, *1*, 87–92.

Lerman, K., Jones, C., Galstyan, A., & Mataric, M. J. (2006). Analysis of dynamic task allocation in multi-robot systems. *The International Journal of Robotics Research*, *25*, 225–242. doi:10.1177/0278364906063426

Li, Q. H. (2002). The application accuracy of the NeuroMate robot - A quantitative comparison with frameless and frame-based surgical localization systems. *International Journal of Computer Aided Surgery*, *7*, 90–98.

Mataric, M. J., Eriksson, J., & Feil-Seifer, D. (2007). Socially assistive robotics for post-stroke rehabilitation. *International Journal of NeuroEngineering and Rehabilitation*, *4*, 1–12. doi:10.1186/1743-0003-4-1

Matsui, T. (2005). *The roles and method of robot design* (pp. 47–55). Proceeding of Robot Design Forum.

McComb, G. (2003). *Robot builder's sourcebook* (pp. 435–440). McGraw-Hill.

McConnell, P., Schneeberger, E., & Michler, R. (2003). History and development of robotic cardiac surgery. *Problems in General Surgery*, *20*, 20–30. doi:10.1097/01.sgs.0000081182.03671.6e

Mellinger, D., Michael, N., & Kumar, V. (2010). Trajectory generation and control for precise aggressive maneuvers with quadrotors. *Proceedings of International Symposium on Experimental Robotics,* (pp. 2520–2525).

Melvin, W., Krause, K., Needleman, B., & Ellison, E. (2003). Robotic resection of a pancreatic neuroendocrine tumor. *Journal of Laparoendoscopic & Advanced Surgical Techniques. Part A.*, *13*, 33–36. doi:10.1089/109264203321235449

Melvin, W., Needleman, B., Krause, K., Schneider, C., & Ellison, E. (2002). Computer–enhanced versus standard laparoscopic antireflux surgery. *Journal of Gastrointestinal Surgery*, *6*, 11–16. doi:10.1016/S1091-255X(01)00032-4

Menzel, P., & D'Aluisio, F. (2000). *Robo Sapiens* (1st ed., pp. 180–189). MIT Press.

Passerotti, C., & Peters, C. (2006). Robotic-assisted laparoscopy applied to reconstructive surgeries in children. *World Journal of Urology*, *24*, 193–197. doi:10.1007/s00345-006-0084-0

Patel, M., Menon, M., & Rogers, C. (2011). Robotic urologic surgery: Robot-assisted partial nephrectomy. *Robotics in Genitourinary Surgery*, *4*, 423–433. doi:10.1007/978-1-84882-114-9_37

Peters, C. (2007). *Applications of robotics in pediatric urologic surgery* (pp. 179–187). Robotic Urologic Surgery. doi:10.1007/978-1-84628-704-6_24

Pettersson, P. O., & Doherty, P. (2004). Probabilistic roadmap based path planning for autonomous unmanned aerial vehicles. *Proceedings of the 14th International Workshop on Connecting Planning and Theory with Practice* (pp. 158–162).

Pott, P., Scharf, H., & Schwarz, M. (2005). Today's state of the art of surgical robotics. *Journal of Computer Aided Surgery*, *10*, 101–132.

Pounds, P., Mahony, R., Gresham, J., Corke, P., & Roberts, J. (2004). Towards dynamically-favorable quad-rotor aerial robots. *Australasian Conference on Robotics and Automation* (pp. 1102 – 1106).

Puntambekar, S., Agarwai, G., Joshi, S., Rayate, N., & Sathe, R. (2009). Robotic radical hysterectomy: Applying principles of the laparoscopic Pune technique. *Journal of Robotic Surgery*, *4*, 259–264. doi:10.1007/s11701-010-0224-9

Rosen, J., Hannaford, B., & Satava, R. (2011). *Surgical robotics - Systems applications and visions*. Springer. ISBN: 978-14419-4

Schiff, J., Li, P. S., & Goldstein, M. (2005). Robotic microsurgical vasovasostomy and epididymostomy in rats. *International Journal of Medical Robotics and Computer Assisted Surgery*, *1*, 122–126. doi:10.1002/rcs.21

Schreuder, H., Aweemer, R., Baal, W., Lande, J., & Dijkstra, J. (2010). From open radical hysterectomy to robot-assisted laparoscopic radical hysterectomy for early stage cervical cancer: Aspects of a single institution learning curve. *Gynecological Surgery*, *7*, 253–258. doi:10.1007/s10397-010-0572-5

Senapati, S., & Advincula, A. (2011). Surgical techniques: Robot-assisted laparoscopic myomectomy with the da Vinci surgical system. *Journal of Robotic Surgery*, *1*, 69–74. doi:10.1007/s11701-007-0014-1

Shakernia, O., Vidal, R., Sharp, C. S., Ma, Y., & Sastry, S. (2002). Multiple view motion estimation and control for landing an unmanned aerial vehicle. *Proceedings of IEEE International Conference on Robotics and Automation,* (pp. 2793–2798).

Sicitiano, B. (1999). Kinematic control of redundant robot manipulators. *Journal of Intelligent & Robotic Systems*, *3*, 201–212. doi:10.1007/BF00126069

Siebert, W., Mai, S., Kober, R., & Heeckt, P. (2004). Total knee replacement: Robotic assistive technique. In DiGioia, A. M., Jaramaz, B., Picard, F., & Nolte, L.-P. (Eds.), *Computer and robotic assisted hip and knee surgery* (pp. 127–156). Oxford University.

Takahashi, T. (2007). *Emotional robot design* (pp. 32–34). Digest Book of IEEE RO-MAN.

Talamini, M., Chapman, W., Horgan, S., & Melvin, W. (2003). Evaluation of 211 robotic surgical procedures. *Surgical Endoscopy*, *17*, 1521–1524. doi:10.1007/s00464-002-8853-3

Tapus, A., Cristian, T., & Mataric, M. J. (2008). User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Journal of Robotics*, *5*, 169–183. doi:10.1007/s11370-008-0017-4

Tapus, A., Matarić, M. J., & Scassellati, B. (2007). The grand challenges in socially assistive robotics. *IEEE Robotics & Automation Magazine*, *14*, 15–18. doi:10.1109/MRA.2007.339605

Taylor, R., & Stoianovici, D. (2003). Medical robotics in computer-integrated surgery. *IEEE Transactions on Robotics and Automation*, 19.

Tholey, G., Chanthasopeephan, T., & Hu, T. (2003). Measuring grasping and cutting forces for reality-based haptic modeling - Computer assisted radiology and surgery. *17th International Congress and Exhibition,* (pp. 158–162).

Tinelli, R., Malzoni, M., Cosentino, F., Perone, C., & Fusco, A. (2011)... *Annals of Surgical Oncology*, *14*, 2654–2661.

Torbjorn, S. D., Mataric, M. J., & Gaurav, S. S. (2009). Multi-robot task allocation through vacancy chain scheduling. *Journal of Robotics and Autonomous Systems*, *57*, 87–91.

Tsagarakis, N., & Caldwell, D. (2003). Development and control of a "soft-actuated" exoskeleton for use in physiotherapy and training. *Autonomous Robots*, *15*, 21–33. doi:10.1023/A:1024484615192

Ullman, D. G. (2002). *The mechanical design process* (pp. 178–182). McGraw-Hill Science.

Yang, C., Huang, Q., & Han, J. (2011). Decoupling control for spatial six-degree-of-freedom electro-hydraulic parallel robot. *Journal of Robotics and Computer-Integrated Manufacturing*, *27*, 1010–1015.

Yang, S. H., Byun, J. H., & Seo, J. H. (2008). The design of kids robot: raBie. *Journal of Korean Society of Design Science*, *21*, 79–82.

Yun, Y., & Li, Y. (2011). Optimal design of a 3-PUPU parallel robot with compliant hinges for micromanipulation in a cubic workspace. *Journal of Robotics and Computer-Integrated Manufacturing*, *27*, 977–985. doi:10.1016/j.rcim.2011.05.001

This work was previously published in Prototyping of Robotic Systems: Applications of Design and Implementation, edited by Tarek Sobh and Xingguo Xiong, pp. 356-378, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 107
Software Process Lines:
A Step towards Software Industrialization

Mahmood Niazi
Keele University, UK & King Fahd University of Petroleum and Minerals, Saudi Arabia

Sami Zahran
Process Improvement Consultant, UK

ABSTRACT

The industrial revolution transformed the cottage industry into mass production. In this chapter, the authors trace the recent advancement of the software industry and establish that it is following a similar route followed by the manufacturing industry towards industrialization. The chapter positions the concept of Software Product Lines (http://www.sei.cmu.edu/productlines/), as a possible foundation for software industrialization, and the authors introduce the concept of Software Process Lines as complimentary foundation for software industrialization. The chapter discusses a number of questions: What are the Software Process Lines? What are the justifications and benefits of Software Process Lines? What are the steps for implementing Software Process Lines? How can Software Process Lines enable and facilitate the establishment of a continuous Software Process Improvement environment?

INTRODUCTION AND JUSTIFICATION

Mega Waves in the Software Industry

The software industry has moved a long way since the early times of it being a cottage industry. In its short history it has passed through a number of mega changes and waves of changes moving firmly towards a discipline similar to what happened to manufacturing through the industrial revolution. The authors believe that the era of software industrialization is coming, in spite of some difficulties facing software development efforts. The first wave of the software industry started with the structured methods in the 1970's, which were developed to cater for the increasing demands and complexity of software. The efforts to adopt structured methods were assisted by automated

DOI: 10.4018/978-1-4666-1945-6.ch107

tools, and this was the beginning of transforming the software industry from a 'cottage industry' towards a disciplined production culture. The second wave of the software industry came with the rise of the object-oriented technology and its associated reuse, combined with the emergence of the focus of the software process maturity and improvement started by Watts Humphrey (Humphrey, 1989) and championed by the Software Engineering Institute (SEI). The software process movement had great promises towards injecting discipline and continuous improvement into the software industry. However this was not enough to move the software industry into the state of industrialization.

Research and industry survey show that the effort put into Software Process Improvement (SPI) models and frameworks can assist in producing high quality software, reducing cost and time, and increasing productivity and employee satisfaction (Ashrafi, 2003; Butler, 1995; Jiang, Klein, Hwang, Huang, & Hung, 2004; Niazi, Wilson, & Zowghi, 2006b; Pitterman, 2000; Yamamura, 1999; Zahran, 1998). However, only a small proportion of the software developing organisations' population has adopted SPI models and framework. This is depicted in the SEI CMMI appraisal report in which only 3113 CMMI appraisals were reported to the SEI (SEI, 2008). Research also shows that a large majority of software development organisations appear to be unwilling to follow SPI initiatives based on process capability maturity models like CMMI (Staples et al., 2007). This may be due to the fact that an SPI programme is quite an expensive undertaking as organisations need to commit significant resources over a long time period (Coleman & O'Connor, 2008). Even organisations who are willing to commit the resources and time do not always achieve their desired results from their SPI initiatives (Florence, 2001; Kautz & Nielsen, 2000). Hence, the significant investment required and limited success are considered two of the main reasons why many organisations are reluctant to embark on a long path of systematic process improvement (Staples et al., 2007). In the case of small and medium size organisations, there are more concerns about the relevance and applicability of SPI models like CMM or CMMI (Brodman & Johnson, 1994; Staples et al., 2007).

Due to complex nature of SPI initiatives it has been proven to be hard for companies to successfully design and implement software process improvement. A typical approach to SPI in general has three stages: process appraisal, process definition/redesign and process implementation and deployment. The process appraisal phase consumes a larger percent of the budget and resources; process definition requires model knowledge, process design knowledge and skills and knowledge of the organisation/company; and often process implementation and deployment is not only across multi-project, but also across multi-site and multi-customer type. The whole SPI initiative is a long-term approach and it takes time to fully implement a SPI initiative. This is one of the reasons that many organisations are reluctant to embark on a long path of systematic process improvement.

Extensive literature review revealed that many standards and models exist for SPI but little attention has been paid to their effective adoption and implementation of these models. The chaotic implementation process is the most common cause of SPI implementation failure (Zahran, 1998). Attention to a defined SPI implementation process is essential for the success of any SPI initiative. In order to address the SPI implementation issues a Software Process Lines approach has been proposed in this chapter. The objective of the Software Process Lines is to reuse the artefacts in SPI initiatives which are costly to develop from scratch. This will not only help organisations to save their time for process definition/ deployment but will also assist them in quickly moving from lower maturity level to higher maturity levels. We believe that the concept of Software Process Lines, introduced in this chapter, in association with the concept of Software Product Line, championed by the SEI (http://www.softwareproductlines.com/), has the potential to instil process discipline and continuous improvement into software development. We believe that are witnessing the early stage of moving towards an era of software industrialization.

Need to Address both the Software Product and the Software Process

In our view, the move towards Software Industrialization requires parallel progress on two fronts: the first is the software product development and the second is the software process maturity and improvement. There are some serious progress and promising results on the front of software product development as reported by the Software Engineering Institute's (SEI) program on Software Product Lines (http://www.sei.cmu.edu/productlines/index.cfm). There should be a similar progress on the front of software development process.

For the last decade there have been many advances in standards and models to improve the software development process. Research shows that the effort put into these SPI models and standards can assist in producing high quality software, reducing cost and time, and increasing productivity (Ashrafi, 2003; Butler, 1995; Pitterman, 2000; Yamamura, 1999). However, these advances have not been matched by equal advances in the adoption of these standards and models in software development (Leung, 1999) which has resulted in limited success for many SPI efforts. Studies show that 67% of SPI managers want guidance on 'how' to implement SPI activities, rather than 'what' SPI activities to actually implement (Herbsleb & Goldenson, 1996). Despite the importance of the SPI implementation process, little empirical research has been carried out on developing ways in which to effectively implement SPI programmes. This suggests that the current problem with SPI is not due to a lack of a standard or model, but rather a lack of effective strategies to successfully implement these standards or models.

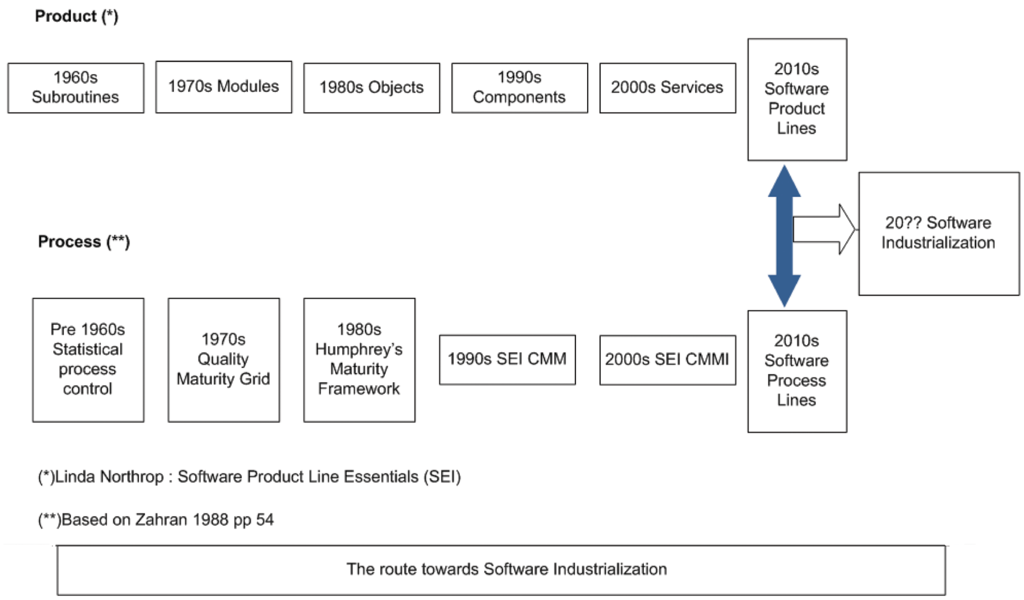
Software quality has received much attention in both academia and industry. This is due to the role software plays in modern-day business and, to some extent, modern-day living. Attempts to improve software quality have been going on for several decades. Software organisations have been struggling with a questionable quality image for their products for a long time (Zahran, 1998). Customer satisfaction has also become the motto of many software organisations in order to survive with quality software (Paulk, Weber, Curtis, & Chrissis, 1994). Software quality has become more critical as software pervades our day-to-day lives (Pitterman, 2000; Yamamura, 1999). We believe that Software Industrialization will provide a way to deliver quality software within budget and schedule, something that currently eludes most software organisations (Ashrafi, 2003; Jiang et al., 2004; Paulk et al., 1994; Standish-Group, 2003). In order to effectively address the software quality issues different approaches have been developed, of which software process improvement (SPI) is the one most widely used (SEI, 2004).

TOWARDS SOFTWARE INDUSTRIALIZATION

We believe that the Software Industry has made big steps over the past five decades towards achieving moving the software industry away from being a "cottage industry" towards achieving a level of process discipline, automation and mass production similar to those achieved in the manufacturing industry as a result of the industrial revolution.

Achieving the status of Industrialization, there is a need to make progress at two parallel fronts: one on the Product Development front, and the other on the Process and Tools front. Figure 1 highlights a number of major milestones of the software industry's progress along the route to-

Figure 1. The route towards software industrialization (Northrop, 2008; Zahran, 2008)



wards Software Industrialization on the two fronts of Product and Process.

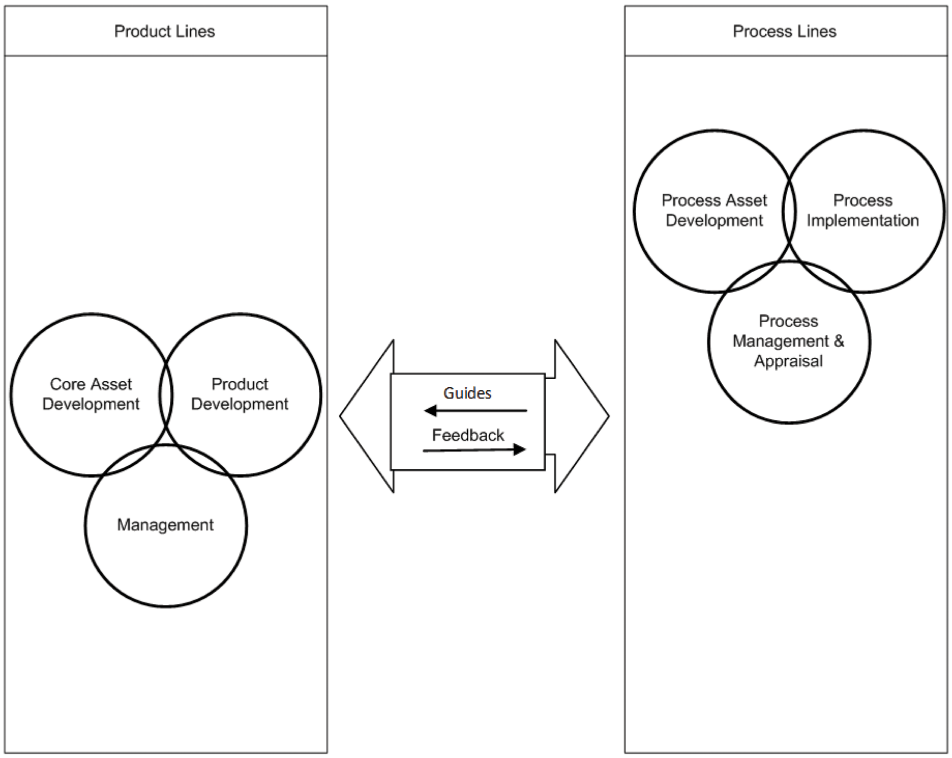
Software Product Lines need to be supported by corresponding Software Process Lines in order to achieve the qualitative shift towards the industrialization of software product development. Successful implementation of Software Product Lines will not be possible in the absence of mature disciplined processes that are continuously optimised and improving. In the case of the manufacturing industry, there was no way for it to achieve its current advanced status without well defined and mostly automated processes. The concept of Software Product Lines is well developed and is now pioneered by a number of product domains in diverse areas including: mobile phones, satellite ground station systems, command and control ship systems, mass storage systems, billing systems, printers and medical devices). Leading organisations who piloted the Software Product Lines and achieved gains and benefits include: Boeing, ERICSSON, NOKIA, HP, ABB, and GM. (Northrop, 2008).

The concept of Software Process Lines, which we developed and describe in this chapter, is still in the early stages of development. Figure 2 illustrates how Product Lines and Process Lines may interact with each other. As illustrated, the Software Process Lines are used to tailor generic processes to generate specific processes to be followed to develop specific products using the generic assets in the Software Product Lines. That means that the Process Lines provide guidance on how to customise and use the core component assets in the Product Lines to develop specific products. In the other direction, there should be feedback from the Product Lines to the Process Lines on the process performance and any need to upgrade or modify the processes in the Process Lines. That feedback should include all the lessons learned resulting from the tailoring of the product lines and any additions to the product lines.

OVERVIEW OF SOFTWARE PRODUCT LINES

Software product lines are emerging as a new and important software development paradigm. A software product line is "a set of software-intensive

Figure 2. Interaction between process lines and product line



systems sharing a common set of features that satisfy the specific needs of a particular market segment or mission" (Clements & Northrop, 1999)

The rationale behind software product lines is that "the practice of building sets of related systems from common assets can yield remarkable quantitative improvements in productivity, time to market, product quality, and customer satisfaction" (Clements & Northrop, 1999) in contrast to being developed one at a time in separate efforts. Using common sets, a new product is formed by taking applicable components from the asset base, tailoring them as necessary through pre-planned variation mechanisms such as parameterisation, adding new components that may be necessary, and assembling the collection under the umbrella of a common, product line-wide architecture (Clements & Northrop, 1999). Building a new product becomes more a matter of generation than creation; the predominant activity is integration rather than programming.

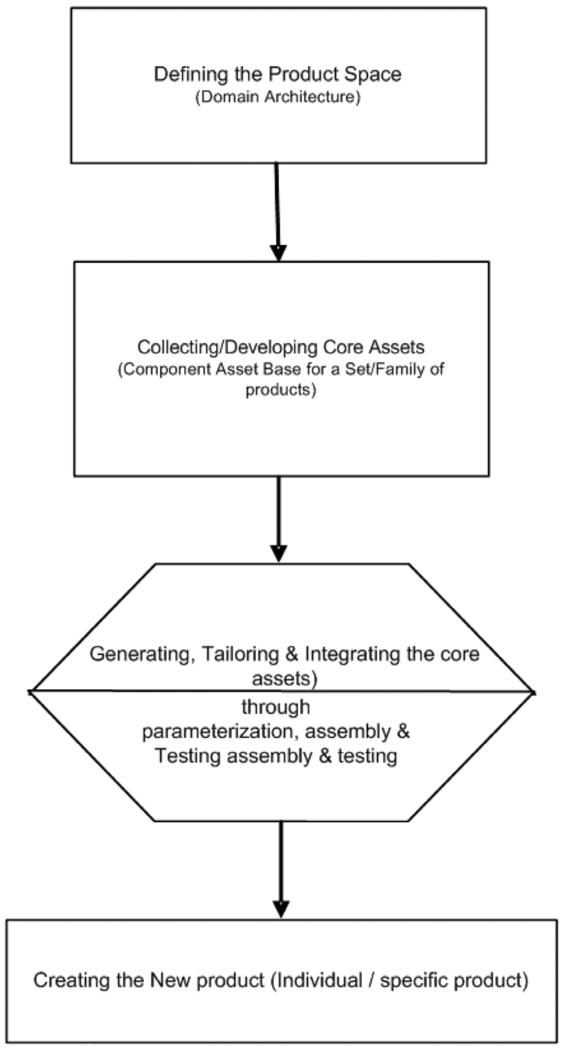
According to the SEI (http://www.sei.cmu.edu/productlines/), organisations of all types and sizes have discovered that a product line strategy, when skilfully implemented, can produce many benefits—and ultimately give the organisations a competitive edge. Example organisational benefits include:

- Improved productivity by as much as 10x
- Increased quality by as much as 10x
- Decreased cost by as much as 60%
- Decreased labour needs by as much as 87%
- Decreased time to market (to field, to launch) by as much as 98%
- Ability to move into new markets in months, not years

Figure 3 below illustrates the main activity flow in a software product line environment.

Generally, adopting a software product line approach involves two classes of activities: Core

Figure 3. Main activity flow in Software Product Lines environment



asset development or acquisition, and Product development or acquisition using core assets. The activities involved are:

- Defining the product space
- Producing the core assets
- Requirements for the specific product
- Developing the production plan

A number of practice areas relevant to each of these activities are suggested by the SEI Product Lines Program in a framework for software product line practice (www.sei.cmu.edu/productlines/frame_report/PL.essential.act.htm)

SOFTWARE PROCESS LINES

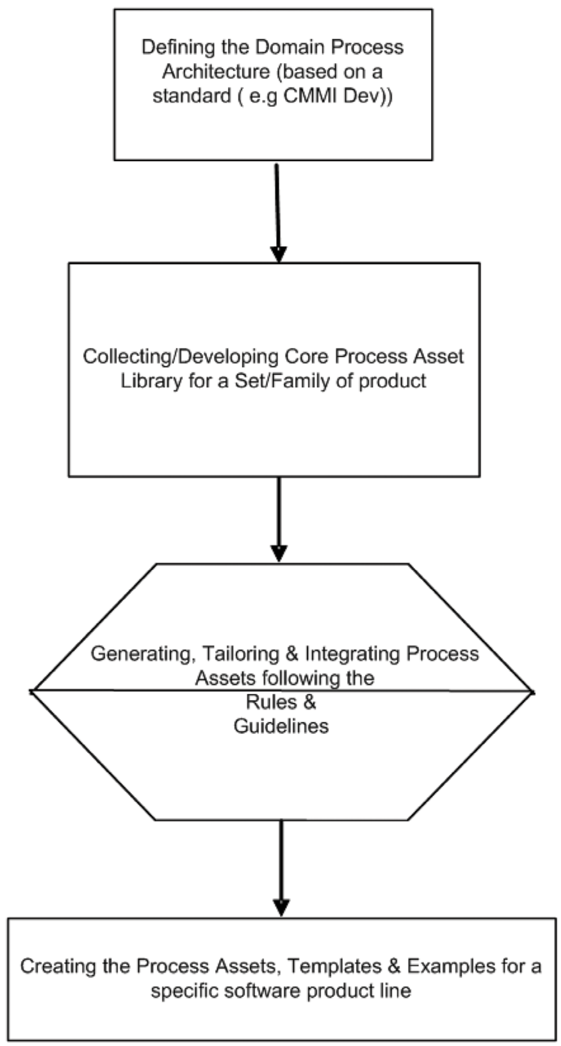
Our suggested Software Process lines approach is based on the concept of software product line approach (Northrop, 2002). As mentioned above, "A software product line is a set of software-intensive systems that share a common, managed feature set satisfying a particular market segment's specific needs or mission and that are developed from a common set of core assets in a prescribed way" (Northrop, 2002). One of the important characteristics of a software product line is Strategic Reuse of a set of reusable product components which has a significant impact on cost of the product.

We have noticed through our experience and working in academia and industry that a similar approach can be adapted for a software process lines which encourages the reuse of generic process assets, thus avoiding the costly effort of developing the process artefacts from scratch. Software Process Line can be defined as "a set of core processes that share a common feature set to satisfy the needs and requirements for process implementation strategies". These core processes form the basis for the software process lines (Figure 4).

Like the essential product line activities described in (Northrop, 2002), we have proposed three essential activities in software process line as shown in Figure 5. Software Process line involves "core process development", "process implementation" and "process management and appraisal".

In the "core process development" activity, generic process patterns will be developed, using process definitions, process patterns, styles and frameworks, which can be adapted by the product development team to their process needs and requirements. Several examples of process pat-

Figure 4. Main activity flow in software process lines environment



terns are already available (Forster, 2006; Telesko & Ferber, 2007; Zahran, 2008). Reuse of these patterns in the process development should save the organisation effort and time in the core process development.

In the "core process implementation" activity, generic process implementation strategies will be developed, using change management methods and process improvement implementation theories, practices and frameworks, which can be adapted by organisations in their SPI implementation initiatives in order to save organisations time for designing new SPI implementation strategies.

The objective of "process management and appraisal" activity is to provide technical and organisational support to other two activities of the software process line in order to successfully implement the software process line initiative.

Core Process Development

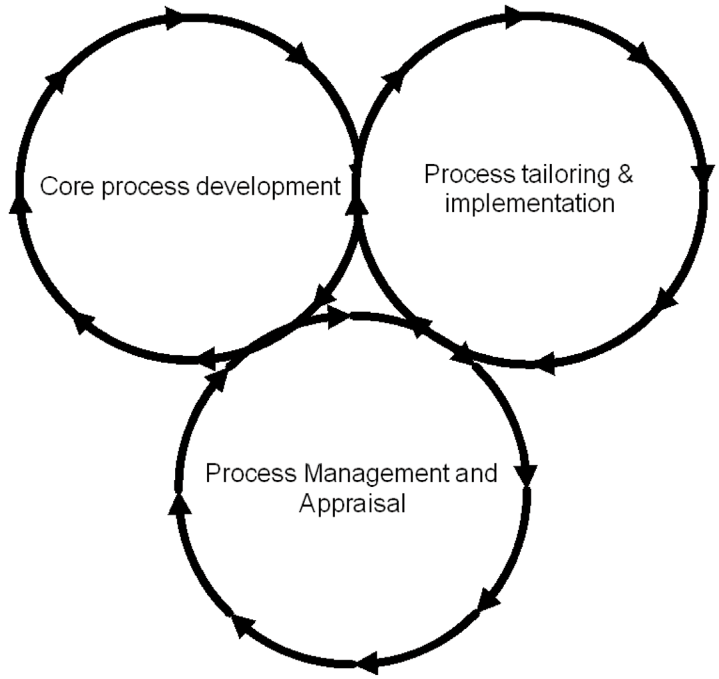
The objective of core process development is to develop core generic process patterns which can easily be adapted by organisations to their SPI needs and requirements as shown in Figure 6. This will not only help organisations to save their time for process definition and development but will also assist them in quickly moving from lower maturity level to higher maturity levels. Research shows that organisations are spending a lot of time in moving from lower maturity level to higher maturity levels which is depicted in the Software Engineering Institute report (SEI, 2004):

- Maturity level 1 to 2 is 22 months
- Maturity level 2 to 3 is 19 months
- Maturity level 3 to 4 is 25 months
- Maturity level 4 to 5 is 13 months

The basic reason for this delay is that in any SPI initiative organisations develop processes from scratch because there are little benchmark processes and process patterns available which might be used to develop these new processes quickly. Despite the importance of this issue, little research has been carried out on software process implementation strategies in general and designing of benchmark process or process patterns in particular. As mentioned earlier only recently some pioneering work on software process patterns has taken place (Forster, 2006; Telesko & Ferber, 2007; Zahran, 2008).

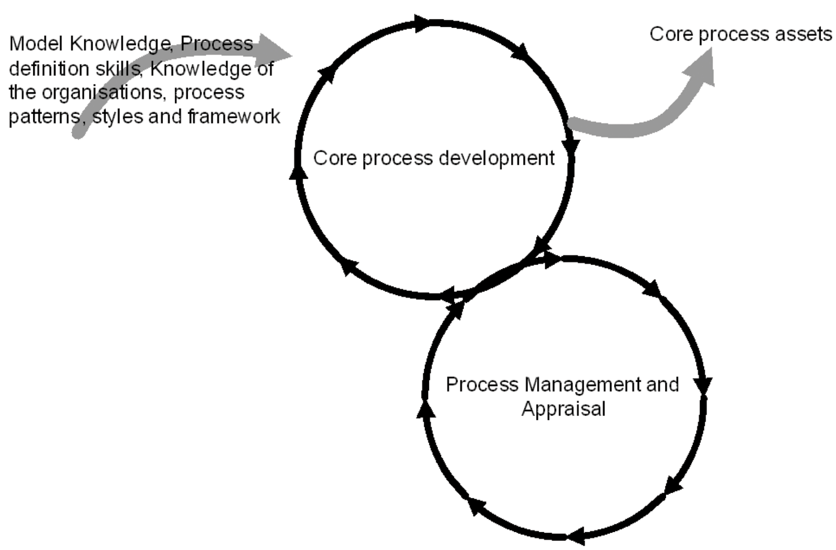
In our previous research (Niazi, Hickman, Ahmad, & Babar, 2008) we have taken initiatives to start work in this missing area of SPI.

Figure 5. Software process lines activities



The objective was, as a pilot project, to design a process pattern for CMMI Level 2 specific practice–"manage requirements changes" as shown in Figure 7. We have proposed a pattern in the form of a model for requirements change management and also discussed initial validation of that model in (Niazi et al., 2008). That model was based on both an empirical study that

Figure 6. Core process line development



we have carried out and our extensive literature review of SPI and requirements engineering. For data collection we have interviewed SPI experts from reputed organisations. Further work included analysing research articles, published experience reports and case studies. Our model is based on five core elements: request, validate, implement, verify and update. Within each of these elements we have identified specific activities that need to take place during requirements change management process.

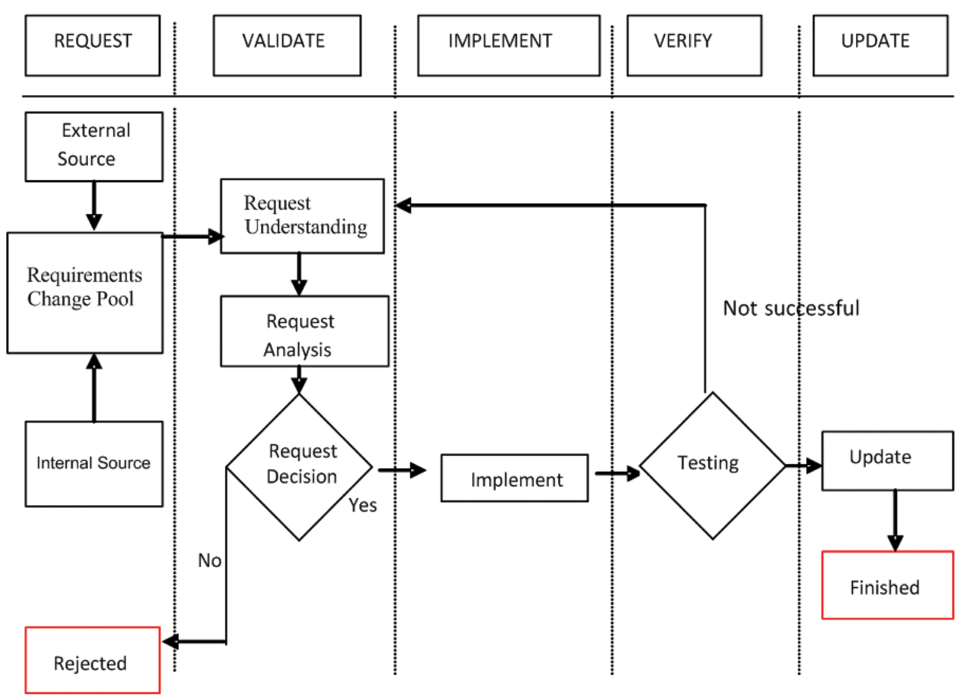
The initial stage is the change "Request". The main sources of requests may be either internal or external. The internal requests come from the project management or software maintenance teams within the company. The external requests come from the customers. These internal and external requests are then fed to a requirements change pool. The requirements change pool contains a description of the change, the reasons behind the changes and who has requested the change.

The next stage is to "Validate" the change request. The first activity in the validate stage is to understand the change request (i.e. what needs to be done–fix, an enhancement or removal). Request analysis is the activity to look at the different ways in which the request can be met, i.e. how much effort is needed to make this change, how much effort is needed to implement this change, the impact of the change, the risk of change and the priority of each change request.

The final activity of the validation process is to make decision if the change request should be accepted, rejected or reanalysed with new evidence. The third stage is to "Implement" the changes. In this stage all the accepted changes are consolidated and are implemented in the form of end product or software.

The fourth stage is to "Verify" changes where it is ascertained that the software conforms to its specification. In this stage the new product/software is tested in the form of regression testing, field trials or user acceptance. The testing method

Figure 7. An example of a process pattern for core process line



will depend on the characteristics of the request and the environment of the product/software. If the verification stage is not successful the change request is sent back to the "Validate" stage for further understanding and analysis of the change request.

The final stage is "Update". Documentation on the software is updated with the changes made. The customers and the project team are all informed about these update so that everyone is working on the current up to date version. The finished step is when the product/software is released with the new requirements included.

The initial evaluation of the model was performed via an expert review process. The initial evaluation of the model showed that the requirements change management model is clear, easy to use and can effectively manage the requirements change process. However, more case studies are needed to evaluate this model in order to further evaluate its effectiveness in the domain of the requirements process.

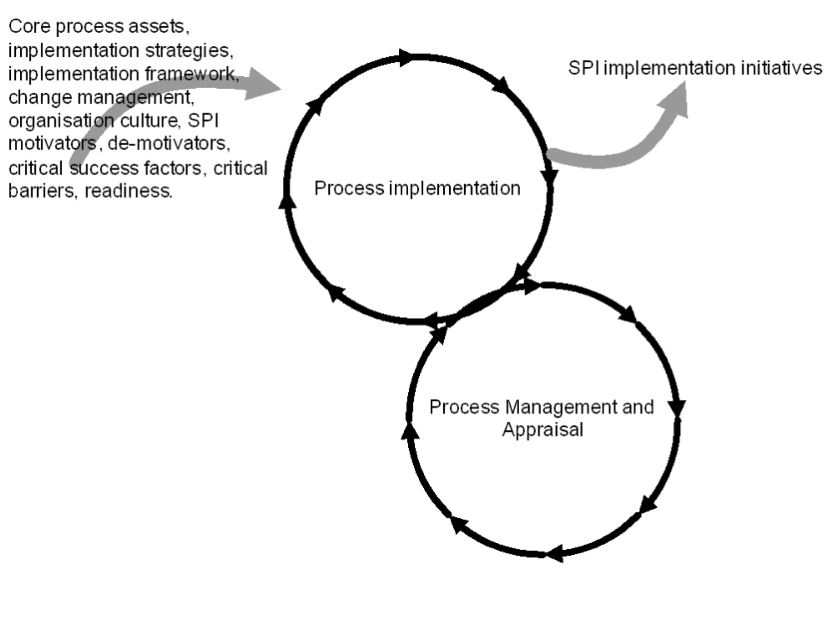
After successfully developing the pattern for one specific practice we are now planning to develop patterns for all of the specific practices of CMMI. However, we encourage other people to conduct more studies to develop new process patterns which can satisfy CMMI specific and generic practices.

Process Implementation

The objective of process implementation activity is to develop process implementation strategies which can easily be adapted by organisations in their software process implementation and process implementation initiatives as shown in Figure 8 This will also help organisations to save time and effort in designing new process implementation and process improvement strategies.

A number of researchers have tried to address some of the issues of SPI implementation by exploring SPI implementation success factors, barriers, motivators and de-motivators (Baddoo & Hall, 2002, 2003; El-Emam, Fusaro, & Smith, 1999; Goldenson & Herbsleb, 1995; Niazi & Babar, 2008, 2007a, 2007b; Niazi & Staples, 2005; Niazi, Wilson, & Zowghi, 2004, 2006a). These studies indeed contribute towards SPI implementation where SPI managers can use their lessons

Figure 8. Process implementation



and concepts as motivators in their SPI implementation strategies. In addition the SPI managers can also design some strategies to cope with and overcome any barriers and or de-motivators to software process improvements in their organisations. However, more work is needed to assist SPI managers in the form of models and frameworks that could help in successfully implementing SPI initiatives. More work is also needed to develop generic implementation strategies which can easily be adapted by organisations in their SPI implementation initiatives to save their time for designing new SPI implementation strategies.

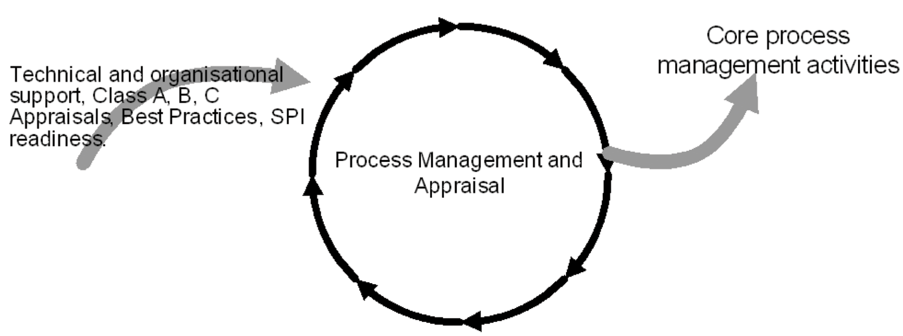
Process Management and Appraisal

The objective of process management and appraisal activity is to provide technical and organisational support to other two activities of process line SPI in order to successfully develop the whole process line SPI initiative as shown in Figure 9. In addition, different appraisals are performed to assess the core process development and process implementation activities. The technical management oversees the core process development and process implementation, ensuring that the stakeholders involved in building processes and patterns engage in the right activities and collect sufficient data to perform these activities. The organisational support ensures that the SPI teams receive the right and enough resources (i.e. experienced people, training, and budget) for SPI activities. In the organisational support the management assist SPI teams to address any barriers or risks that threaten Software Process Line activities. Finally, someone should be designated as the process line manager to lead all the Software Process Line activities.

In order to perform process appraisals, there are three classes available of CMMI SCAMPI appraisals Class A, B, and C (Ahern et al., 2005; CMMI-Product-Team, 2001). Class A appraisals are very costly, time-consuming, and resource-intensive, but provide the highest levels of assurance about their findings. Class B appraisals are less costly, time-consuming and resource-intensive as they use fewer appraisers, appraise fewer projects, and rely on fewer forms of evidence. Class C appraisals are the cheapest and easiest to perform, and can approach the simplicity of a structured questionnaire. All appraisals result in a report of findings about an organisation's capability, but only Class A appraisals can result in a publicly-reportable "rating" of the organisation's CMMI level.

Conducting research on SPI implementation has convinced us to conduct some research on appraisal of organisations readiness for SPI implementation. It is very important to measure organisations readiness for SPI implementation as the failure rate for SPI programmes has been reported up to 70% in a report from the Software Engineering Institute (SEI) (Ngwenyama

Figure 9. Process management and appraisal



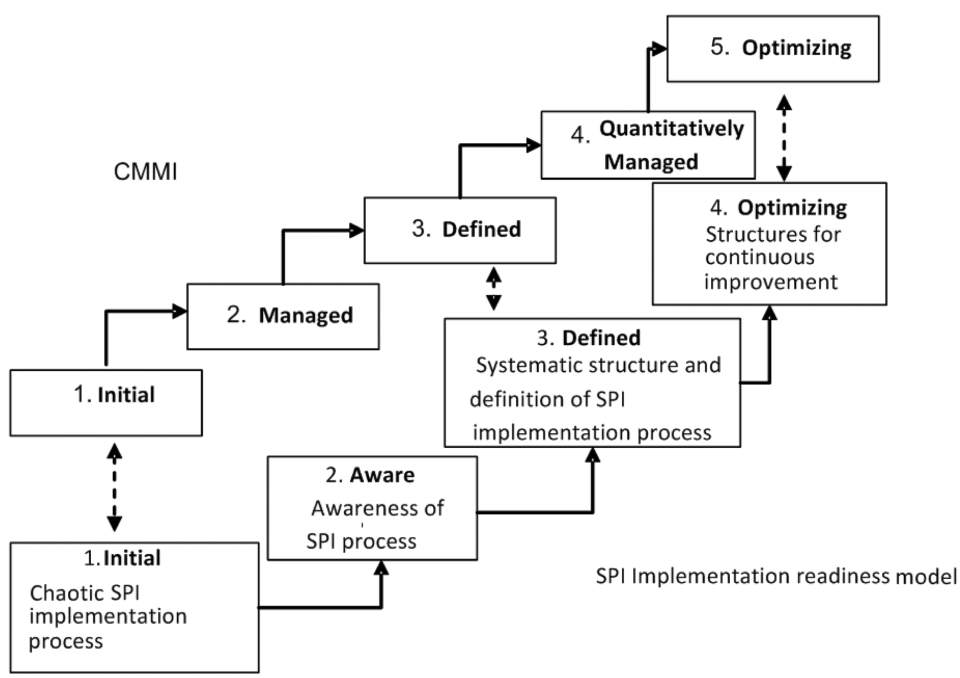
& Nielsen, 2003; SEI, 2002). This shows that many organisations go for SPI implementation prior to testing their readiness for SPI activities. We have focused on these issues and developed a SPI readiness model (as shown in Figure 10) in order to assess the SPI implementation readiness of the organisations (Niazi et al., 2005). The CMMI perspective (Chrissis, Konrad, & Shrum, 2006) and the findings from our previous empirical study (Niazi, Wilson, & Zowghi, 2003; Niazi et al., 2004) were used in the design of the SPI readiness model. The SPI readiness model has four SPI implementation readiness levels abstracted from CMMI. These readiness levels contain different critical success factors (Niazi et al., 2006a) and critical barriers (Niazi et al., 2004) identified through the literature and interviews. Under each factor, different practices have been designed that guide how to assess and implement each factor. We have evaluated this model with three large sized and two medium sized organisations. The evaluation results showed that this model can be used effectively to identify SPI implementation issues with a goal of increasing SPI implementation readiness. All the case study participants agreed that this model is clear, easy to use and specifically geared to assess the organisations' SPI implementation readiness.

CONCLUSION AND FUTURE WORK

In conclusion, here is a summary of the main points in this chapter:

- Software process improvement is still a costly and complex endeavour for many organisations.
- There is a lack of re-use of software process patterns to create core software processes and instil software process improvement practices in the organisation's fabric.
- The SEI Software Product Lines initiative is achieving success and resulting in

Figure 10. SPI implementation readiness model (Niazi, Wilson, & Zowghi, 2005)



benefits for organisations that adopted this approach.

- We believe that an effort to develop Software Process Lines is necessary to complement, support and enhance the value of adopting the Software Product Lines approach.
- In this chapter we have introduced our approach to Software Process Lines, which we believe together with the Software Product Lines can provide a good springboard to moving towards the Software Industrialization

We suggest that further work is still needed in progressing our concepts forward, especially in the following areas which we recommend them for students and researchers:

- Prototyping our approach to Software Process Lines in order to prove the concepts viability in commercial and business environment
- Piloting the combined implementation of the Software Product Lines and the Software Process Lines (in a software factory-like environment) to order to prove the concept of Software Industrialization
- Elaborating of the definitions and descriptions of the basic concepts described in this chapter and formulating them in coherent methodology for developing and using Software Process Lines
- Practicing and piloting the Software Process Lines activity flow shown in Figure 3, and elaborating the activities and the flow based on the lessons learned resulting from the pilots.
- Suggesting more details of the possible interface and interaction between Process Lines and Products Lines as outlined in Figure 2.

The authors welcome any initiative in the above direction and are willing to provide advice and support to any effort in this direction.

REFERENCES

Ahern, D. M., Armstrong, J., Clouse, A., Ferguson, J. R., Hayes, W., & Nidiffer, K. E. (2005). *CMMI SCAMPI distilled: Appraisals for process improvement*. Addison-Wesley Professional.

Ashrafi, N. (2003). The impact of software process improvement on quality: In theory and practice. *Information & Management*, *40*(7), 677–690. doi:10.1016/S0378-7206(02)00096-4

Baddoo, N., & Hall, T. (2002). Motivators of software process improvement: An analysis of practitioner’s views. *Journal of Systems and Software*, (62): 85–96. doi:10.1016/S0164-1212(01)00125-X

Baddoo, N., & Hall, T. (2003). De-motivators of software process improvement: An analysis of practitioner’s views. *Journal of Systems and Software*, *66*(1), 23–33.

Brodman, J. G., & Johnson, D. L. (1994). What small businesses and small organizations say about the CMMI. In *Proceedings of 16th International Conference on Software Engineering* (pp. 331-340). IEEE Computer Society Press.

Butler, K. (1995). The economics benefits of software process improvement. *CrossTalk*. *The Journal of Defense Software Engineering*, *8*(7).

Chrissis, M. B., Konrad, M., & Shrum, S. (2006). *CMMI guidelines for process integration and product improvement*. Addison Wesley.

Clements, P., & Northrop, L. (1999). *A framework for software product line practice*. Retrieved Dec 2009, from http://www.sei.cmu.edu/library/abstracts/news-at-sei/backgroundsep99pdf.cfm

CMMI-Product-Team. (2001). *Appraisal requirements for CMMI, version 1.1.* Technical Report CMU/SEI-2001-TR-034, Carnegie Mellon University Software Engineering Institute.

Coleman, G., & O'Connor, R. (2008). Investigating software process in practice: A grounded theory perspective. *Journal of Systems and Software*, *81*, 772–784. doi:10.1016/j.jss.2007.07.027

El-Emam, K., Fusaro, P., & Smith, B. (1999). Success factors and barriers for software process improvement. In Messnarz, R., & Tully, C. (Eds.), *Better software practice for business benefit: Principles and experience* (pp. 355–371). Los Alamitos, CA: IEEE Computer Society.

Florence, A. (2001). Lessons learned in attempting to achieve software CMM Level 4. *CrossTalk, The Journal of Defense Software Engineering*, 29-30.

Forster, F. (2006). *Business process improvement patterns.* Doctoral dissertation, Queensland University of Technology & Technische Universitat, Munich.

Goldenson, D. R., & Herbsleb, J. D. (1995). *After the appraisal: A systematic survey of process improvement, its benefits, and factors that influence success* (No. CMU/SEI-95-TR-009, Software Engineering Institute, USA): SEI.

Herbsleb, J. D., & Goldenson, D. R. (1996). *A systematic survey of CMM experience and results.* Paper presented at the 18th International Conference on Software Engineering (ICSE-18), Berlin, Germany.

Humphrey, W. (1989). *Managing the software process*. Addison-Wesley.

Jiang, J., Klein, G., Hwang, H., Huang, J., & Hung, S. (2004). An exploration of the relationship between software development process maturity and project performance. *Information & Management*, *41*, 279–288. doi:10.1016/S0378-7206(03)00052-1

Kautz, K., & Nielsen, P. A. (2000). *Implementing software process improvement: Two cases of technology transfer.* Paper presented at the 33rd Hawaii Conference on System Sciences, Maui, USA.

Leung, H. (1999). Slow change of information system development practice. *Software Quality Journal*, *8*(3), 197–210. doi:10.1023/A:1008915509865

Ngwenyama, O., & Nielsen, P., A. (2003). Competing values in software process improvement: An assumption analysis of CMM from an organizational culture perspective. *IEEE Transactions on Engineering Management*, *50*(1), 100–112. doi:10.1109/TEM.2002.808267

Niazi, M., & Babar, M. A. (2007a). *De-motivators for software process improvement: An analysis of Vietnamese practitioners' views.* Paper presented at the International Conference on Product Focused Software Process Improvement PROFES 2007, LNCS 4589.

Niazi, M., & Babar, M. A. (2007b). *Motivators of software process improvement: An analysis of vietnamese practitioners' views.* Paper presented at the International Conference on Evaluation and Assessment in Software Engineering (EASE 2007).

Niazi, M., & Babar, M. A. (2008). Demotivators for software process improvement: An empirical investigation. *Software Process Improvement and Practice Journal (Perspectives on Global Software Development: Special Issue on PROFES 2007), 13*(3), 249-264.

Niazi, M., Hickman, C., Ahmad, R., & Babar, M. A. (2008). *A model for requirements change management: Implementation of CMMI level 2 specific practice.* Paper presented at the International Conference on Product Focused Software Process Improvement PROFES 2008, Italy, LNCS 5089.

Niazi, M., & Staples, M. (2005). *Systematic review of organizational motivations for adopting CMM-based SPI.* Technical Report (No. National ICT Australia, PA005957).

Niazi, M., Wilson, D., & Zowghi, D. (2003). *Critical success factors and critical barriers for software process improvement: An analysis of literature*. Paper presented at the Australasian Conference on Information Systems (ACIS03), Perth, Australia.

Niazi, M., Wilson, D., & Zowghi, D. (2004). *Critical barriers for SPI implementation: An empirical study.* Paper presented at the IASTED International Conference on Software Engineering (SE 2004), Austria.

Niazi, M., Wilson, D., & Zowghi, D. (2005). A maturity model for the implementation of software process improvement: An empirical study. *Journal of Systems and Software*, *74*(2), 155–172. doi:10.1016/j.jss.2003.10.017

Niazi, M., Wilson, D., & Zowghi, D. (2006a). Critical success factors for software process improvement: An empirical study. *Software Process Improvement and Practice Journal*, *11*(2), 193–211. doi:10.1002/spip.261

Niazi, M., Wilson, D., & Zowghi, D. (2006b). *Implementing software process improvement initiatives: An empirical study.* Paper presented at the The 7th International Conference on Product Focused Software Process Improvement, LNCS.

Northrop, L. (2002). SEI's software product line tenets. *IEEE Software*, *19*(4), 32–40. doi:10.1109/MS.2002.1020285

Northrop, L. (2008). Software product line essential. Retrieved Dec 2009, from http://www.sei.cmu.edu/library/assets/spl-essentials.pdf

Paulk, M., Weber, C., Curtis, B., & Chrissis, M. (1994). *A high maturity example: Space shuttle onboard software, in the capability maturity model: Guidelines for improving software process*. California, US: Addison-Wesley.

Pitterman, B. (2000). Telcordia Technologies: The journey to high maturity. *IEEE Software*, *17*(4), 89–96. doi:10.1109/52.854074

SEI. (2002). *Process maturity profile of the software community*. Software Engineering Institute Carnegie Mellon University.

SEI. (2004). *Process maturity profile*. Software Engineering Institute Carnegie Mellon University.

SEI. (2008). *Process maturity profile*. Software Engineering Institute Carnegie Mellon University. Retrieved Oct 2008, from http://www.sei.cmu.edu/appraisal-program/profile/pdf/CMMI/2008MarCMMI.pdf

Standish-Group. (2003). *Chaos - The state of the software industry*.

Staples, M., Niazi, M., Jeffery, R., Abrahams, A., Byatt, P., & Murphy, R. (2007). An exploratory study of why organizations do not adopt CMMI. *Journal of Systems and Software*, *80*(6), 883–895. doi:10.1016/j.jss.2006.09.008

Telesko, R., & Ferber, S. (2007). *Applying design patterns on processes.* Paper presented at the SEI SEPG European Conference.

Yamamura, G. (1999). Software process satisfied employees. *IEEE Software*, (September/October): 83–85. doi:10.1109/52.795105

Zahran, S. (1998). *Software process improvement - Practical guidelines for business success*. Addison-Wesley.

Zahran, S. (2008). *Patterns for the enterprise process architecture*. Paper presented at the SEI SEPG European Conference.

This work was previously published in Software Process Improvement and Management: Approaches and Tools for Practical Development, edited by Shukor Sanim Mohd Fauzi, Mohd Hairul Nizam Md Nasir, Nuraminah Ramli and Shamsul Sahibuddin, pp. 1-17, copyright 2012 by Information Science Reference (an imprint of IGI Global).

Chapter 108
Super High Efficiency Multi-Junction Solar Cells and Concentrator Solar Cells

Masafumi Yamaguchi
Toyota Technological Institute, Japan

ABSTRACT

While single-junction solar cells may be capable of attaining AM1.5 efficiencies of up to 29%, Multi-Junction (MJ, Tandem) III-V compound solar cells appear capable of realistic efficiencies of up to 50% and are promising for space and terrestrial applications. In fact, the InGaP/GaAs/Ge triple-junction solar cells have been widely used in space since 1997. In addition, industrialization of concentrator solar cell modules using III-V compound MJ solar cells have been announced by some companies. This chapter presents principles and key issues for realizing high-efficiency MJ solar cells, issues relating to development and manufacturing, and applications for space and terrestrial uses.

INTRODUCTION

Multi-Junction (MJ, Tandem) solar cells are composed of multi-layers with different bandgap energies are shown in Figure 1 and have the potential for achieving high conversion efficiencies of over 50% and are promising for space and terrestrial applications due to wide photo response. Figure 2 shows theoretical conversion efficiencies of single-junction and Multi-Junction (MJ) solar cells in comparison with experimentally realized efficiencies.

Tandem solar cells were proposed by Jackson (1955) and Wolf (1960). Table 1 shows progress of the III-V compound multi-junction solar cell technologies. MIT group (Fan, Tsaur, & Palm, 1982) encouraged R&D of tandem cells based on their computer analysis. Although AlGaAs/GaAs tandem cells, including tunnel junctions and metal interconnectors, were developed in the early years, a high efficiency close to 20% was not obtained (Hutchby, Markunas, & Bedair,

DOI: 10.4018/978-1-4666-1945-6.ch108

Figure 1. A schematic structure of a multi-layer solar cell

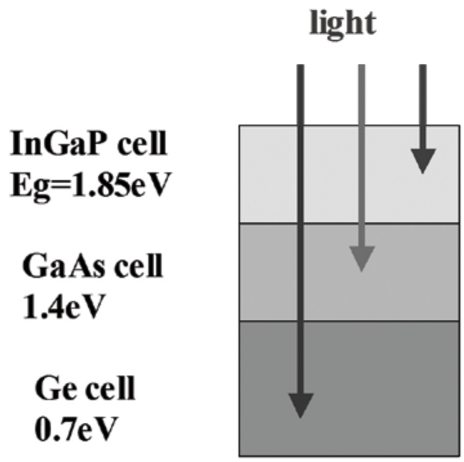
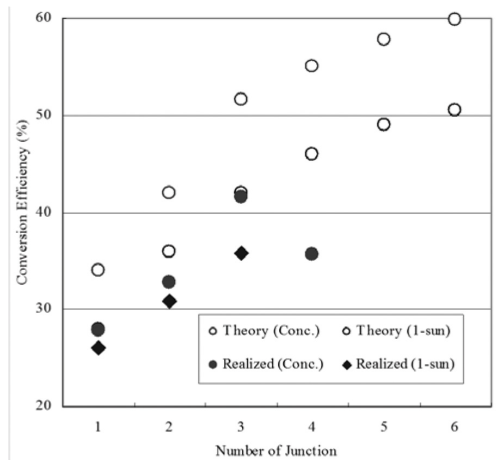


Figure 2. Theoretical conversion efficiencies of single-junction and multi-junction solar cells in comparison with experimentally realized efficiencies



1985). This is because of difficulties in making high performance and stable tunnel junctions, and the defects related to the oxygen in the AlGaAs materials (Ando, Amano, Sugiura, et al., 1987). A Double Hetero (DH) structure tunnel junction was found to be useful for preventing diffusion from the tunnel junction and improving the tunnel junction performance by the authors (Sugiura, Amano, Yamamoto, & Yamaguchi, 1988). The authors demonstrated 20.2% efficiency AlGaAs/GaAs 2-junction cells (Amano, Sugiura, Yamamoto, & Yamaguchi, 1987). An InGaP material for the top cell was proposed by NREL group (Olson, Kurtz, & Kibbler, 1990). As a result of performance improvements in tunnel junction and top cell, over 30% efficiency has been obtained with InGaP/GaAs 2-junction cells by the authors (Takamoto, Ikeda, Kurita, et al., 1997).

InGaP/GaAs-based MJ solar cells have drawn increased attention for space applications because superior radiation-resistance of InGaP top cells and materials have been discovered by the authors (Yamaguchi, Okuda, Taylor, et al., 1997). and those have the possibility of high conversion efficiency of over 30%. In fact, the commercial satellite (HS 601HP) with 2-junction GaInP/GaAs-on Ge solar arrays was launched in 1997 (Brown, Goldhammer, Goodelle, et al., 1997).

More recently, InGaP/GaAs-based MJ solar cells have drawn increased attention for terrestrial applications because concentrator operation of MJ cells have great potential of providing high performance and low-cost solar cell modules. For concentrator applications, grid structure has been designed in order to reduce the energy loss due to series resistance, and 38.9% (AM1.5G, 489-suns) efficiency has been demonstrated by Sharp (Takamoto, Kaneiwa, Imaizumi, & Yamaguchi, 2005). Most recently, 41.6% efficiency has been reported with InGaP/GaAs/Ge 3-junction concentrator cells by Spectrolab (King, Boca, Hong, et al., 2009). In addition, the authors have realized high-efficiency and large-area (5,445cm^2) concentrator InGaP/InGaAs/Ge 3-junction solar cell modules of an outdoor efficiency of 31.5% (Araki, Uozumi, Egami, et al., 2005). as a result of developing high-efficiency InGaP/InGaAs/Ge 3-junction cells, low optical loss Fresnel lens and homogenisers, and designing low thermal conductivity modules. Some companies including Sharp (Tomita, 2006) have announced to commercialise

Table 1. Progress of the III-V compound multi-junction solar cell technologies

1955 1960	Proposal of multi-junction solar cell	Jackson Wolf
1982	Efficiency calculation of tandem cells	MIT
1982	15.1% AlGaAs/GaAs 2-junction (2-J) cell	RTI
1987	Proposal of double-hetero structure tunnel junction for multi-junction interconnection	NTT
1987	20.2% AlGaAs/GaAs 2-J cell	NTT
1989	32.6% GaAs//GaSb concentrator 2-J cell (mechanical-stacked, 100-suns concentration)	Boeing
1990	Proposal of InGaP as top a cell material	NREL
1990	27.3% InGaP/GaAs 2-J cell	NREL
1996	30.3% InGaP/GaAs 2-J cell	Jpn. Energy
1997	Discovery of radiation-resistance of InGaP top cell	Toyota Tech. Inst.
1997	33.3% InGaP/GaAs//InGaAs 3-J cell (mechanical-stacked)	Jpn. Energy, Sumitomo & Toyota Tech. Inst.
1997	Commercial satellite with 2-J cells	Hughes
2000	31.7% InGaP/InGaAs/Ge 3-J cell	Jpn. Energy
2003	37.4% InGaP/InGaAs/Ge 3-J cell (200-suns concentration)	Sharp
2004	38.9% InGaP/InGaAs/Ge 3-J cell (489-suns concentration)	Sharp & Toyota TI
2006	31.5% large-area (5,445cm^2) InGaP/InGaAs/Ge 3-J cell module (outdoor)	Daido Steel, Daido Metal, Sharp &Toyota T.I.
2006	40.7% InGaP/GaAs/Ge 3-J cell (236-suns concentration)	Spectrolab
2009	41.1% InGaP/InGaAs/Ge 3-J cell (454-suns concentration)	Fraunhofer ISE
2009	41.6% InGaP/InGaAs/Ge 3-J cell (364-suns concentration)	Spectrolab
2009	35.8% InGaP/GaAs/InGaAs 3-J cell (1-sun)	Sharp

InGaP/GaAs/Ge 3-junction concentrator cell modules for terrestrial use.

KEY ISSUES FOR REALIZING HIGH-EFFICIENCY MJ SOLAR CELLS

Key issues for realizing high-efficiency MJ tandem cells are discussed based on our results.

Selection of Cell Materials and Improving the Quality

MJ cells with different band gaps are stacked in tandem so that the cells cover wide wavelength region from 300 nm to 1800 nm. Cell materials are selected by considering band gap energies close to the optimal band gap energy combination based on theoretical efficiency calculation, and by considering lattice matching to substrates and less impurity problems. Figure 3 shows minority-carrier diffusion length dependence of GaAs single-junction solar cell efficiency. It is clear that the higher minority-carrier diffusion length L (minority-carrier lifetime $\tau = L^2 / D$, where D is minority-carrier diffusion coefficient) is substantially necessary to realize the higher efficiency solar cells. Figure 4 shows carrier concentration dependence of minority-carrier lifetime in p-type and n-type GaAs (Ahrenkiel, Keyes,

Figure 3. Minority-carrier diffusion length dependence of GaAs single-junction solar cell efficiency

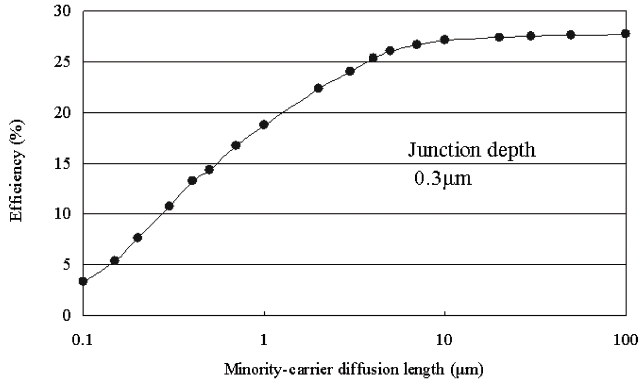
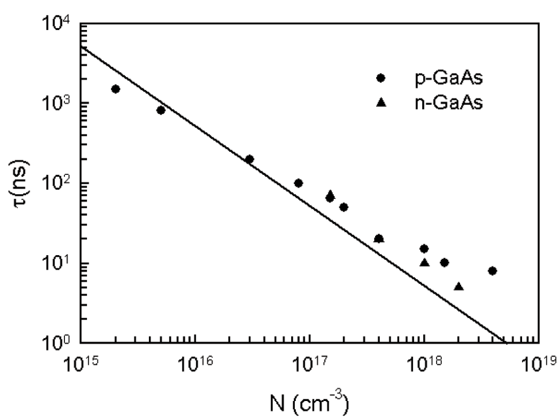


Figure 4. Carrier concentration dependence of minority-carrier lifetime in p-type and n-type GaAs (Ahrenkiel, Keyes, Durbin, & Gray, 1993)



Durbin, & Gray, 1993). Minority-carrier lifetime τ is dependent on carrier concentration N of solar cell layers as expressed by:

$$\tau = 1 / BN \tag{1}$$

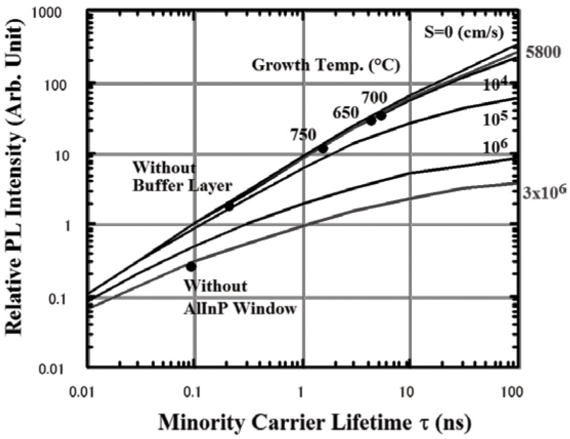
where, B is radiative recombination coefficient. Therefore, carrier concentration of cell layers must be optimised by considering minority-carrier lifetime, build-in potential, and series resistance of p-n junction diodes.

Selection of cell materials, especially selection of top cell materials is also important for high-efficiency tandem cells. It has been found by the authors (Hutchby, Markunas, &. Bedair, 1985) that oxygen-related defect in the AlGaAs top cell materials acts as the recombination centre. As a top cell material latticed matched to GaAs or Ge substrates, InGaP has some advantages (Olson, Kurtz, & Kibbler, 1990) such as lower interface recombination velocity, less oxygen problem and good window layer material compared to AlGaAs. The top cell characteristics depend on the minority carrier lifetime in the top cell layers. Figure 5 shows changes in Photoluminescence (PL) intensity of the solar cell active layer as a function of the minority carrier lifetime τ of the p-InGaP base layer grown by MOCVD (metal-organic chemical vapour deposition) and surface recombination velocity (S). The lowest S was obtained by introducing the AlInP window layer and the highest τ was obtained by introducing buffer layer and optimising the growth temperature. The best conversion efficiency of the InGaP single-junction cell was 18.5% (Yang, Yamaguchi, Takamoto, et al., 1997).

Lattice Matching between Cell Materials and Substrates

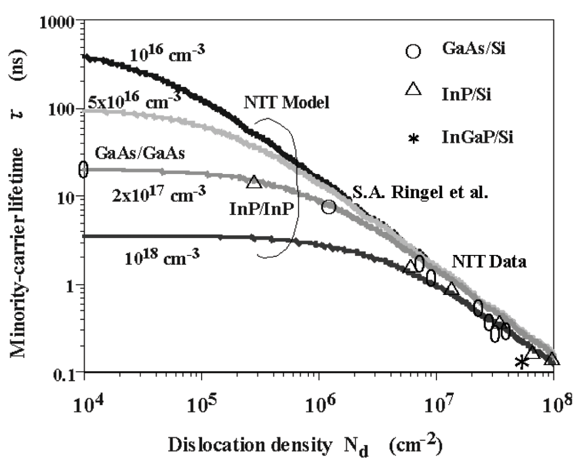
Lattice mismatching of cell materials to substrates should be decreased because miss-fit dislocations must be generated in the upper cell layers and deteriorate cell efficiency. Figure 6 shows calculated and experimental dislocation density depen-

Figure 5. Changes in photoluminescence (PL) intensity of the solar cell active layer as a function of the minority carrier lifetime in GaAs, grown by MOCVD and surface recombination velocity (S)



dence of minority-carrier lifetime in GaAs (Yamaguchi & Amano, 1985). Dislocation density N_d dependence of minority-carrier lifetime τ is expressed by the following equation (Yamaguchi & Amano, 1985):

Figure 6. Calculated and experimental dislocation density dependence of minority-carrier lifetime in GaAs, InP, and InGa



$$1/\tau = 1/\tau_r + 1/\tau_0 + \pi^3 D N_d / 4 \qquad (2)$$

where τ_r is radiative recombination lifetime and τ_0 is minority-carrier lifetime associated with recombination at other unknown defects.

Application of InGaAs middle cell (Yamaguchi & Amano, 1985) lattice-matching to Ge substrates has demonstrated to increase open-circuit voltage (V_{oc}) due to lattice-matching and short-circuit current density (J_{sc}) due to decrease in bandgap energy of middle cell.

Effectiveness of Wide Bandgap Back Surface Field (BSF) Layer

Figure 7 shows surface recombination effect on short-circuit current density J_{sc} of $In_{0.14}Ga_{0.86}As$ homo-junction solar cells as a function of junction depth. Therefore, in order to improve efficiency drop attributed from front and rear surface recombination as shown in Figure 8, formation of heteroface or double-hetero structure is necessary.

Figure 8 shows changes in V_{oc} and J_{sc} of InGaP single-junction cells as a function of potential barrier ΔE. Wide bandgap Back-Surface Field (BSF) layer (Yamaguchi & Amano, 1985) is found

Figure 7. Surface recombination effect on short-circuit current density Jsc of $In_{0.14}Ga_{0.86}As$ homo-junction solar cells as a function of junction depth

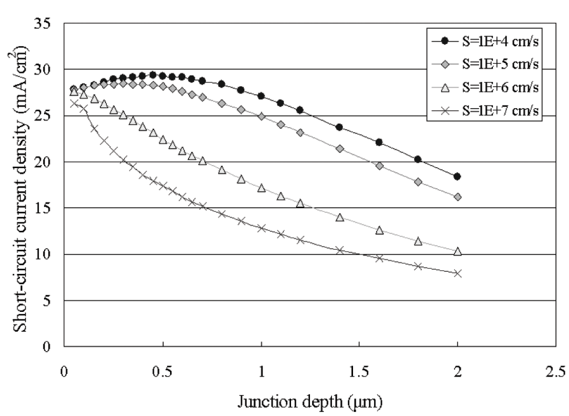
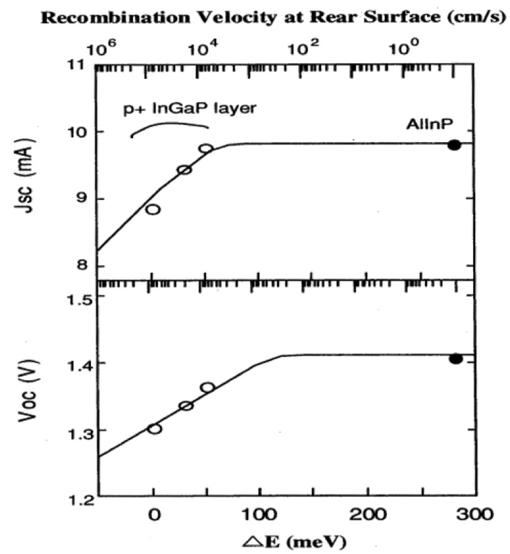


Figure 8. Changes in Voc and Jsc of InGaP single-junction cells as a function of potential barrier ΔE



to more effective for confinement of minority carriers compared to highly doped BSF layers.

Low Loss Tunnel Junction for Intercell Connection and Preventing Impurity Diffusion from Tunnel Junction

Another important issue for realizing high-efficiency monolithic-cascade type tandem cells is the achievement of optically and electrically low-loss interconnection of two or more cells. A degenerately doped tunnel junction is attractive because it only involves one extra step in the growth process. To minimize optical absorption, formation of thin and wide-bandgap tunnel junctions is necessary as shown in Figure 9. However, the formation of a wide-bandgap tunnel junction is very difficult because the tunnelling current decreases exponentially with increase in bandgap energy.

In addition, impurity diffusion from a highly doped tunnel junction during overgrowth of the top cell increases the resistivity of the tunnel junction. As shown in Figure 10, a Double Hetero (DH) structure was found to be useful for preventing diffusion by the authors (Sugiura, Amano, Yamamoto, & Yamaguchi, 1988). An InGaP tunnel junction has been for the first time tried for an InGaP/GaAs tandem cell in our work (Takamoto, Ikeda, Kurita, et al., 1997). As p-type and n-type dopants, Zn and Si were used, respectively. Peak tunnelling current of the InGaP tunnel junction is found to increase from 5mA/cm^2 up to 2A/cm^2 by making a DH structure with AlInP barriers. Therefore, the InGaP tunnel junction has been observed to be very effective for obtaining high tunnelling current, and DH structure has also been confirmed to be useful for preventing diffusion.

DH structure effect on suppression of impurity diffusion from the tunnel junction has been examined. Effective suppression of the Zn diffusion from tunnel junction by the InGaP tunnel junction with the AlInP-DH structure is thought to be attributed to the lower diffusion coefficient (Takamoto, Yamaguchi, Ikeda, et al., 1999) for Zn in the wider bandgap energy materials such as the AlInP barrier layer and InGaP tunnel junction layer.

Table 2 summarizes key issues of realizing super-high-efficiency MJ solar cells

Figure 9. Calculated tunnel peak current density and short-circuit current density Jsc of GaAs bottom cell as a function of bandgap energy of tunnel junction

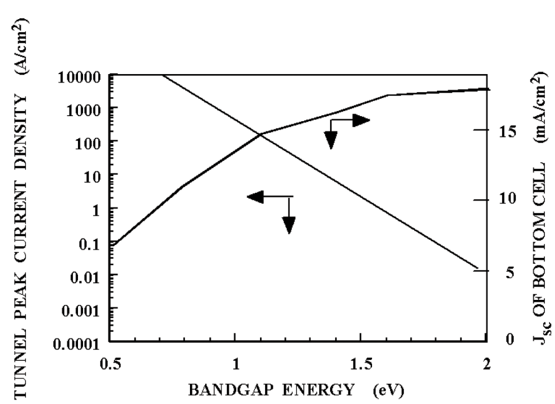
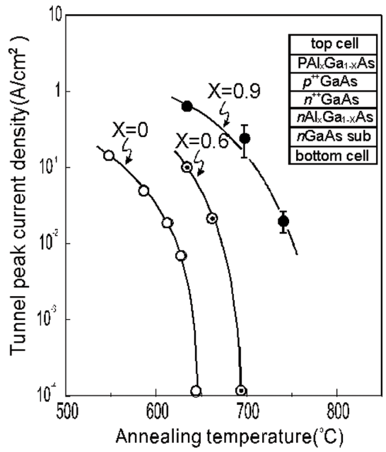


Figure 10. Annealing temperature dependence of tunnel peak current densities for double hetero structure tunnel diodes. X is the Al mole fraction in AlxGa1-xAs barrier layers.



HIGH-EFFICIENCY INGAP/GAAS/GE 3-JUNCTION SOLAR CELLS AND THEIR SPACE APPLICATIONS

Development of High-Efficiency InGaP/GaAs/Ge 3-Junction Solar Cells

As one of the Sunshine Program in Japan, an R&D project for super high-efficiency MJ solar cells was started in 1990. Conversion efficiency of InGaP/GaAs based multijunction solar cells has been improved by the following technologies. A schematic illustration of the InGaP/(In)GaAs/Ge triple junction solar cell and key technologies for improving conversion efficiency are shown in Figure 11.

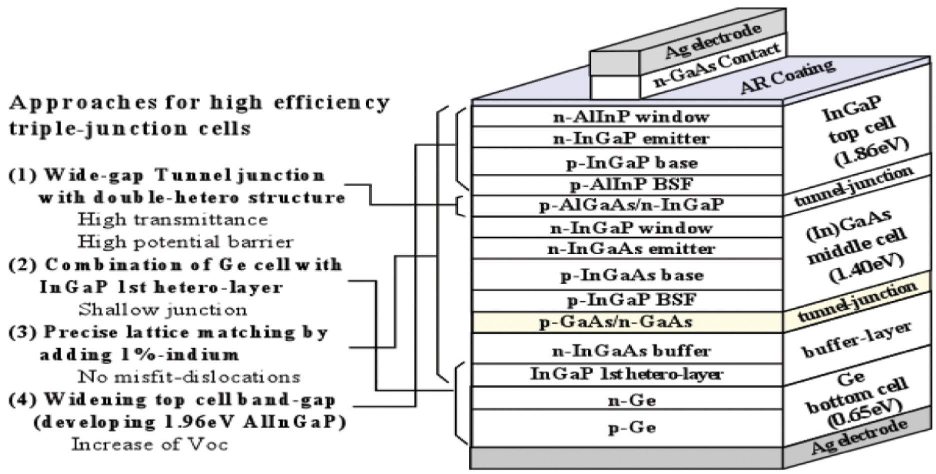
Wide Band-Gap Tunnel Junction

A wide band-gap tunnel junction which consists of double-hetero structure p-Al(Ga)InP/p-AlGaAs/n-(Al)InGaP/n-Al(Ga)InP increases incident light into the (In)GaAs middle cell and produces effective potential barriers for both minority-carriers generated in the top and middle cells. Both Voc and Isc of the cells are improved by the wide band-gap tunnel junction without absorption and recombination losses (Takamoto, Ikeda, Kurita, et al., 1997).. It is difficult to obtain high tunnelling peak current with wide gap tunnel junction, so thinning depletion layer width by formation of highly doped junction is quite necessary. Since impurity diffusion is occurred during growth of the top cell (Sugiura, Amano, Yamamoto, & Yamaguchi, 1988), carbon and silicon which have low diffusion coefficient are used for p-type AlGaAs and n-type (Al)InGaP, respectively. Furthermore,

Table 2. Key issues for realizing super high-efficiency multi-junction solar cells

Key Issue	Past	Present	Future
Top cell materials	AlGaAs	InGaP	AlInGaP
3rd layer materials	None	Ge	InGaAsN etc.
Substrate	GaAs	Ge	Si
Tunnel junction	DH-structure GaAs Tunnel Junction (TJ)	DH-structure InGaP TJ.	DH-structure InGaP or GaAs TJ
Lattice matching	GaAs middle cell	InGaAs middle cell	(In)GaAs middle cell
Carrier confinement	InGaP-BSF	AlInP-BSF	Widegap-BSF Quantum Dots (QDs)
Photon confinement	None	None	Bragg reflector, QDs, etc.
others	None	Inverted Epi.	Inverted Epi. Epitaxial Lift Off

Figure 11. Schematic illustration of a triple-junction cell and approaches for improving efficiency of the cell

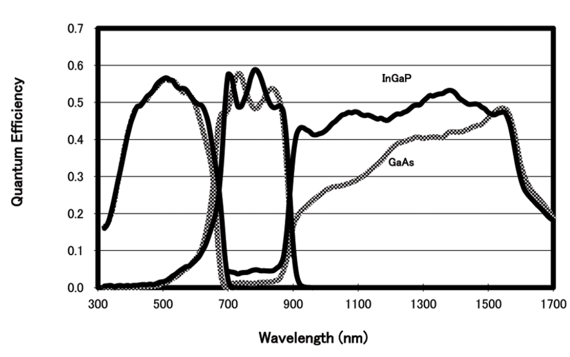


the double-hetero structure supposes to suppress impurity diffusion from the highly doped tunnel junction (Takamoto, et al., 1999). The second tunnel junction between middle and bottom cells consists of p-InGaP/p-(In)GaAs/ n-(In)GaAs/n-InGaP which have wider band-gap than middle cell materials.

Heteroface Structure Ge Bottom Cell

InGaP/GaAs cell layers are grown on a p-type Ge substrate. PN junction is formed automatically during MOCVD growth by diffusion of V-group atom from the first layer grown on the Ge substrate. So, the material of the first hetero layer is important for the performance of Ge bottom cell. An InGaP layer is thought to be suitable material for the first hetero layer, because phosphor has lower diffusion coefficient in Ge than arsenic and indium has lower solubility in Ge than gallium. Figure 12 shows the change in spectral response of the triple-junction cell by changing the first hetero growth layer on Ge from GaAs to InGaP. Quantum efficiency of the Ge bottom cell was improved by the InGaP hetero-growth layer. In the case of GaAs hetero-growth layer, junction depth was measured to be around 1μm. On the other hand, thickness of n-type layer produced by

Figure 12. Change in the spectral response due to modification of the 1st hetero-layer from GaAs to InGaP (without anti-reflection coating)



phosphor from the InGaP layer was 0.1µm. An increase in Ge quantum efficiency was confirmed to be due to a reduction in junction depth.

Precise Lattice-Matching to Ge Substrate

Although 0.08% lattice-mismatch between GaAs and Ge was thought to be negligibly small, misfit-dislocations were generated in thick GaAs layers and deteriorated cell performance. By adding about 1% indium into the InGaP/GaAs cell layers, all cell layers are lattice-matched precisely to the Ge substrate. As a result, crosshatch pattern caused by misfit-dislocations due to lattice-mismatch was disappeared in the surface morphology of the cell with 1% indium, as shown in Figure 13. The misfit-dislocations were found to influence not to Isc but to Voc of the cell. Voc was improved by eliminating misfit-dislocations for the cell with 1% indium. In addition, wavelength of the absorption edge became longer and Isc of both top and middle cells increased, by adding 1% indium.

As one of the Sunshine Program in Japan, an R&D project for super high-efficiency MJ solar cells was started based on our results in 1990. Conversion efficiency of InGaP-based 3-junction solar cells has been improved by the following technologies:

1. Selection and high quality growth of InGaP as a top cell material
2. Proposal of double-hetero structure and wide-band gap tunnel junction for cell interconnection.
3. Precise lattice matching of InGaP top cell and InGaAs middle cell with Ge substrate
4. Proposal of AlInP as a back surface field layer for the InGaP top cell
5. Proposal of InGaP-Ge heteroface structure bottom cell

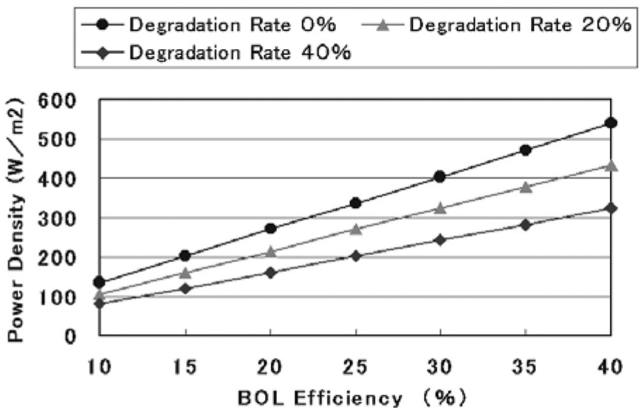
As a result of the above proposals and performance improvements, we have demonstrated a world-record efficiency (33.3% at 1-sun AM1.5G) InGaP/GaAs/InGaAs 3-junction solar cell in 1997 (Takamoto, Ikeda, Agui, et al., 1997). The conversion efficiency of InGaP/(In)GaAs/Ge 3-junction solar cells has been improved to 31.7% (AM1.5G) (Takamoto, et al., 2000).

Radiation-Resistance of InGaP-Based MJ Solar Cells

Figure 14 shows effectiveness of radiation-resistance and high conversion efficiency of space cells from the point view of power density. Since radiation in space is severe, particularly in the Van Allen radiation belt, lattice defects are induced in semiconductors due to high-energy electron and proton irradiations, and the defects cause a decrease in the output power of solar cells. Further improvements in conversion efficiency and radiation-resistance of space cells are necessary for widespread applications of space missions. Recently, InGaP/GaAs-based MJ solar cells have drawn increased attention because of the possibility of high conversion efficiency of over 40% and radiation-resistance. An AM0 efficiency of 29.2%

Figure 13. Surface morphology of InGaAs with various indium compositions grown on Ge

Figure 14. Effectiveness of radiation-resistance and high conversion efficiency of space cells from the point view of power density



has been demonstrated for an InGaP/InGaAs/Ge 3-junction cell ($4cm^2$) (Takamoto, et al., 1997).

Figure 15 shows the maximum power recovery due to light illumination of 100 mW/cm^2 at various temperatures for 1-MeV electron-irradiated InGaP/GaAs tandem cells (Yamaguchi, et al., 1997). The ratios of maximum power after injection, P_I, to maximum power before irradiation, P_0, are shown as a function of injection time. Even at room temperature, photo injection-enhanced annealing of radiation damage to InGaP/GaAs tandem cells was observed. The recovery ratio increases with an increase in ambient temperature within the operating range for space use. Such a recovery is found to be attributed from damage recovery in InGaP top cell layer (Yamaguchi, et al., 1997). Therefore, the results show that InGaP/GaAs tandem cells under device operation conditions have superior radiation-resistant properties.

Figure 15. The maximum power recovery of the InGaP/GaAs tandem cell due to light illumination at various temperatures

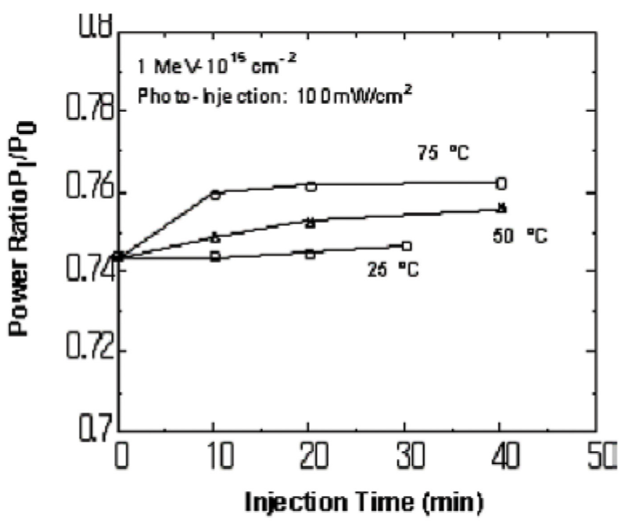
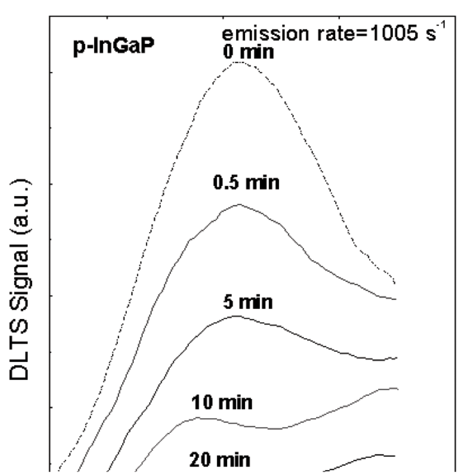


Figure 16 shows DLTS (Deep Level Transient Spectroscopy) spectrum of Trap H2 (Ev+0.55eV) for various injection times at 25^{0}C with an AM1.5 light intensity of 100mA/cm^2. It is also found (Kahn et al., 2000) by DLTS measurements that a major defect level H2 (Ev+0.55eV) recovers by forward bias or light illumination. Moreover, the H2 center is confirmed to act as a recombination center by using the double carrier pulse DLTS method. The enhancement of defect annealing in InGaP top cell layer under minority-carrier injection conditions is thought to occur as a result of the nonradiative electron-hole recombination process (Lang & Leung, 1976) whose energy E_R enhances the defect motion. The thermal activation energy E_A (1.1eV) of the defect is reduced to E_I (0.48~0.54eV) by an amount E_R (0.56~0.62eV). Thus electronic energy from a recombination event can be channeled into the lattice vibration mode which drives the defect motion: $E_I = E_A - E_R$.

Space Applications of InGaP/GaAs/Ge 3-Junction Solar Cells

Advanced technologies for high efficiency cells and discovery of superior radiation-resistance of InGaP based materials are thought to contribute to industrialization of InGaP-based multijunction space solar cells in Japan. Figure 17 shows Sharp space solar cell conversion efficiency heritage. Since 2002, InGaP/GaAs/Ge 3-junction solar cells have been commercialized for space use in Japan.

Figure 16. DLTS spectrum of trap H2 (Ev+0.55eV) for various injection timesat 25^{0}C with an injection density of 100mA/cm^2



LOW COST POTENTIAL OF CONCENTRATOR MJ SOLAR CELL MODULES AND HIGH EFFICIENCY CONCENTRATOR INGAP/GAAS/GE 3-JUNCTION SOLAR CELL MODULES AND TERRESTRIAL APPLICATIONS

Concentrator operation is very effective for cost reduction of solar cell modules and thus that of PV systems. Figure 18 shows configuration of PV system composed of solar cell, optics and tracker. Concentrator operation of the MJ cells is essential for their terrestrial applications. Since the concentrator PV systems using MJ solar cells have great potential of cost reduction as shown in Figure 19 (Yamaguchi, et al., 2003), R&D on concentrator technologies including MJ cells is started in Japan.

In order to apply a high efficiency MJ cell developed for 1 sun condition to a concentrator cell operating under ~500suns condition, reduction in energy loss due to series resistance is the most important issue. Cell size was determined to be 7mm x 7mm with considering total current flow. Grid electrode pitching, height and width were designed in order to reduce series resistance. Figure 20 shows fill factor (FF) of the cell with various grid pitching under 250 suns. Grid electrode with 5μm height and 5μm width was made of Ag. Grid pitching influences lateral resistance between two grids (RL) and total electrodes resistance (RE). Series resistance of the cell (RS), RE and RL are also shown in Figure 20. RE was measured directly after removing electrode from the cell by chemical etching. RL was calculated

Figure 17. Sharp space solar cell conversion efficiency heritage

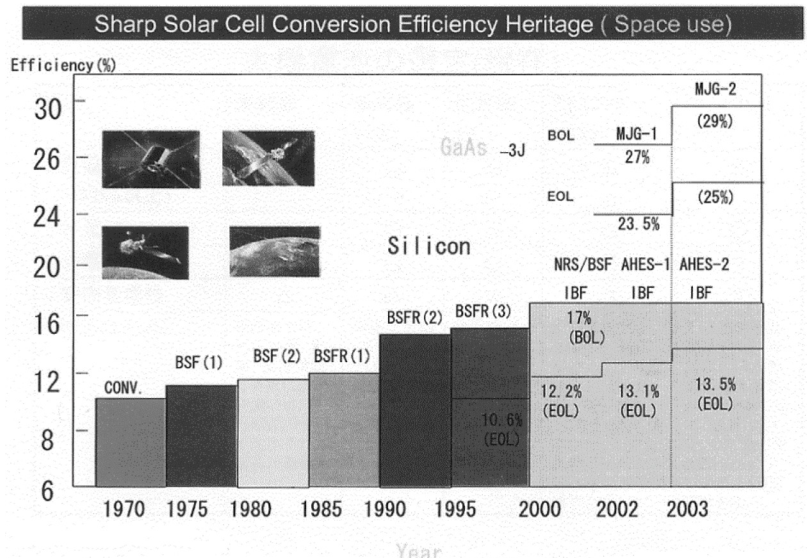


Figure 18. Configuration of PV system composed of solar cell, optics, and tracker

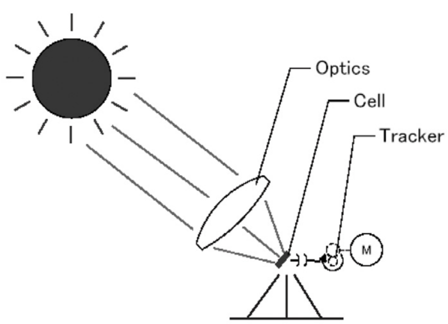
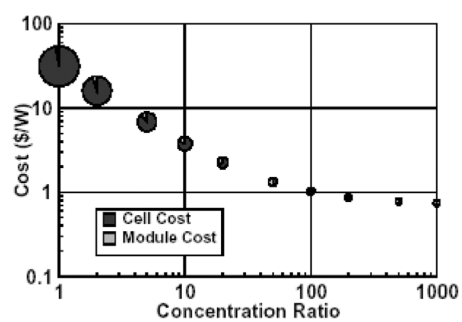


Figure 19. Summary of estimated cost for the concentrator PV systems vs. concentration ratio (Yamaguchi, et al., 2003)



by using sheet resistance of window and emitter layers. Based on the data in Figure 20, the grid pitching is determined to be 0.12 mm at this time. In order to reduce series resistance down to 0.01Ω and obtain high FF under 500 suns, grid height should be increased to be twice. High efficiency under <500 suns is thought to be obtained by the optimal grid design without modification of the cell layer.

For concentrator applications, the grid structure has been designed in order to reduce the energy loss due to series resistance as shown in Figure 20. The authors have successfully fabricated high efficiency concentrator InGaP/InGaAs/Ge 3-junction solar cells designed for 500-sun application. The efficiencies by in-house measurement are 39.2% at 200-suns and 38.9% at 489-suns as shown in Figure 21 (Takamoto, et al., 1997). The solar simulator was equipped with both Xe lamp and halogen lamp and adjusted AM1.5G spectrum. Most recently, 41.6% efficiency has been reported with InGaP/GaAs/Ge 3-junction concentrator cells by Spectrolab (King, et al., 2009).

Figure 20. FF of the concentrator cells with various grids pitching under 250-suns light. Series resistance (Rs), lateral resistance (RL) and total electrodes resistance (RE) are also shown.

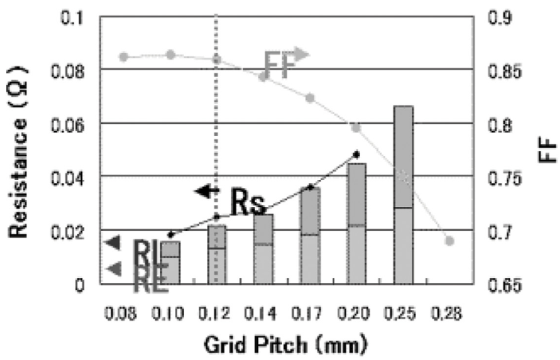
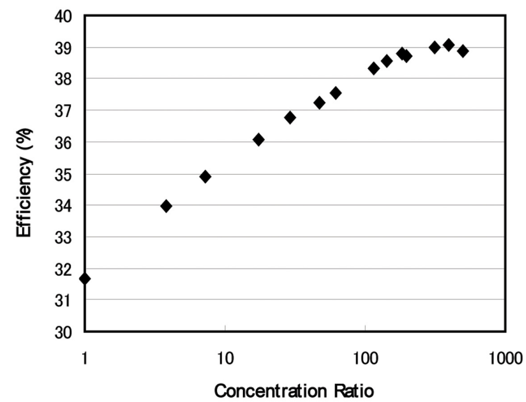


Figure 21. Efficiency of a high-efficiency InGaP/InGaAs/Ge 3-junction cell vs. number of suns



Concentrator InGaP/GaAs/Ge 3-junction solar cell modules have also been developed for terrestrial use (Araki, et al., 2005). A new concentrator optics is introduced, consist of a non-imaging dome-shaped Fresnel lens, and a kaleidoscope homogenizer. The non-imaging Fresnel lens allows wide acceptance half angle with keeping the same optical efficiency with minimum chromatic aberration. The homogenizer reshapes the concentrated into square solar cell aperture, mixed rays to uniform flux. Injection molding is capable of manufacturing thousands of lenses in a single day and by a single machine. The drawback of this method is difficulty of creating precise prism angles and flat facets. The maximum efficiency was a little above 80% and overall efficiency was 73%. After improvement of the process conditions, the averaged efficiency raised to 85.4%.

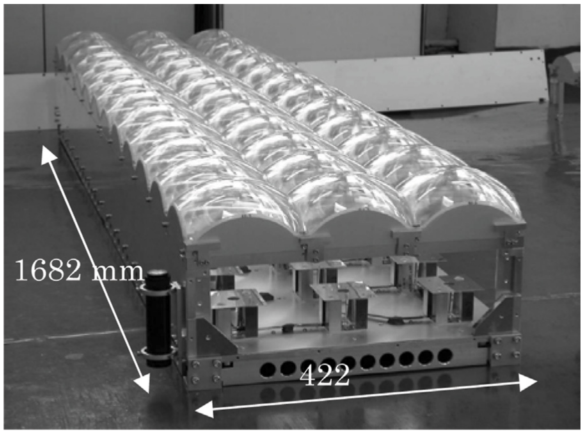
A new packaging structure for III-V concentrator solar cells is developed, applicable mainly to Fresnel lens concentrator modules but may also be used in dish concentrator systems. The solar cell used in the new receiver package is III-V 3-junction concentrator solar cell developed. It is grown on a fragile Ge substrate with thickness of only 150 um. The overall size was 7 mm X 9 mm with 7 mm square aperture area.

In addition, the following technologies have been developed:

1. Super-high pressure and vacuum-free lamination of the solar cell that suppresses the temperature rise to 20 degrees under 550 X geometrical concentration illumination of sunbeam.
2. Direct and voids-free soldering technologies of the fat metal ribbon to the solar cell, suppressing hot-spots and reducing the resistance, thereby allowing a current 400 times higher than normal non-concentration operation to be passed with negligible voltage loss.
3. A new encapsulating polymer that survives exposure to high concentration UV and heat cycles.
4. Beam-shaping technologies that illuminate the square aperture of the solar cell, from a round concentration spot.
5. Homogeniser technologies that give a uniform flux and prevent the conversion losses that stem from chromatic aberration and flux intensity distribution.

The concentrator module is designed with ease of assembly in mind. All the technologically complex components are packaged into a

Figure 22. Inside of the 400 X concentrator module with 36 receivers connected in series



receiver so that a series of receivers and lenses can be assembled with standard tools, using local materials and workforce. The concept is similar to the computer and automobile assembly industries, where key components are imported but the product assembled locally. It is anticipated that this approach will reduce the manufacturing cost of the module as shown in Figure 22.

The peak uncorrected efficiency for the 7,056 cm^2 400 X module with 36 solar cells connected in series was 26.6%, measured in house. The peak uncorrected efficiencies for the same type of module with 6 solar cells connected in series and 1,176 cm^2 area measured by Fraunhofer ISE and NREL were 27.4% and 24.9%. The 5,445 cm^2 550 X modules have also demonstrated 27-28.9%, measured in house. Table 3 summarizes the measured efficiency in three different sites.

A new 400X and 550X (geometrical concentration ratio) are developed and show the highest efficiency in any types of PV as well as more than 20 years of accelerated lifetime. This achievement is blessed with new innovative concentrator technologies. The new concentrator system is expected to open a door to a new age of high efficiency PV.

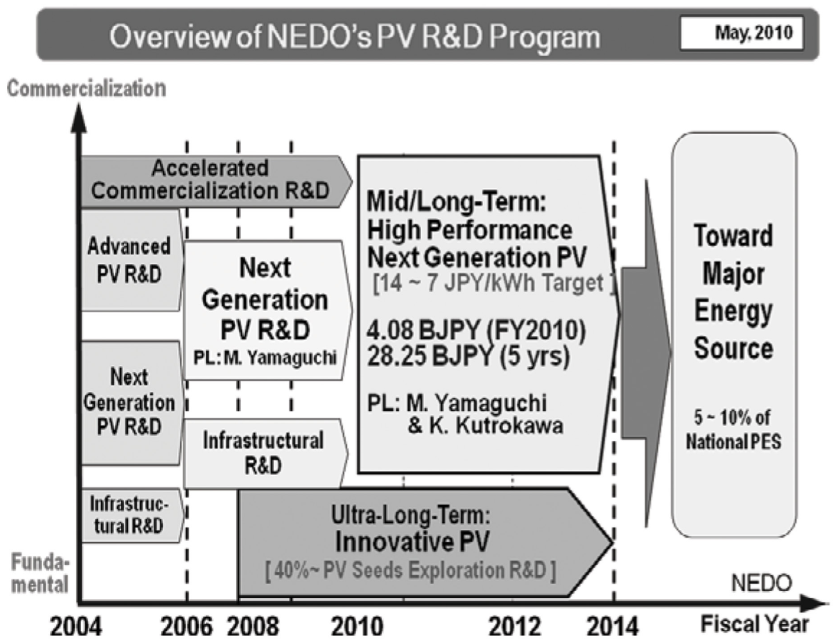
MOST RECENT RESULTS OF MJ CELLS

Recently, more than 40% efficiency cells were reported by Fraunhofer ISE (Bett, et al., 2009) and Spectrolab (King, et al., 2009). Concentrator 4-junction or 5-junction solar cells have great potential for realizing super high efficiency of over 50%. We have been studying concentrator multi-junction solar cells under Japanese Innovative Photovoltaic R&D program started since FY2008 (King, et al., 2009). Figure 23 shows overview of NEDO's PV R&D Program. Japanese Innovative

Table 3. Uncorrected peak efficiency measurement

Concentration	Area cm^2	Site	Ambient	Uncorrected Efficiency	DNI W/m^2
400 X	7,056	Inuyama, Japan Manufacturer	29 C	27.6%	810
400 X	7,056	Toyohashi, Japan Independent	7 C	25.9%	645
400 X	1,176	Fraunhofer ISE, Germany Independent	19 C	27.4%	839
400 X	1,176	NREL, USA Independent	29 C	24.9%	940
550 X	5,445	Inuyama, Japan Manufacturer	33 C	28.9%	741
550 X	5,445	Toyohashi, Japan Independent	28 C	27%	777

Figure 23. Overview of NEDO's PV R&D program



Photovoltaics R&D Program has been started from fiscal year 2008 and the target in this program is to develop high efficiency solar cells with conversion efficiency of more than 40% and low electricity cost of less than 7 JPY/kWh until 2050.

Most recently, world-record efficiency (36.9%) at 1-sunAM1.5G has been realized with inverted epitaxial grown InGaP/GaAs/InGaAs 3-junction cells by Sharp (Takamoto, et al., 2010). Figure 24 fabrication process of InGaP/GaAs/InGaAs 3-J solar cell inverted epitaxial grown. Figure 24 shows I-V curve of world-record efficiency InGaP/GaAs/InGaAs solar cell. Figure 25 shows chronological improvements in conversion efficiencies of III-V compound MJ solar cells under 1-sun and concentrator conditions (see Figure 26).

Figure 24. Fabrication process of InGaP/GaAs/InGaAs 3-J solar cell inverted epitaxial grown

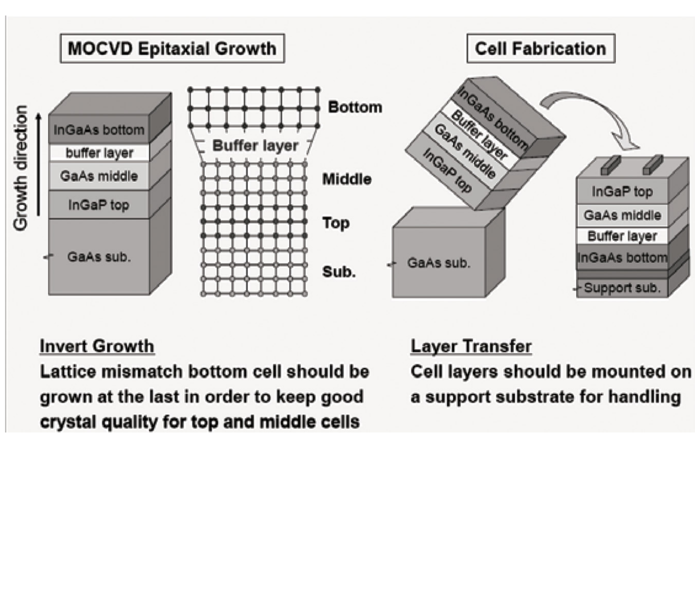
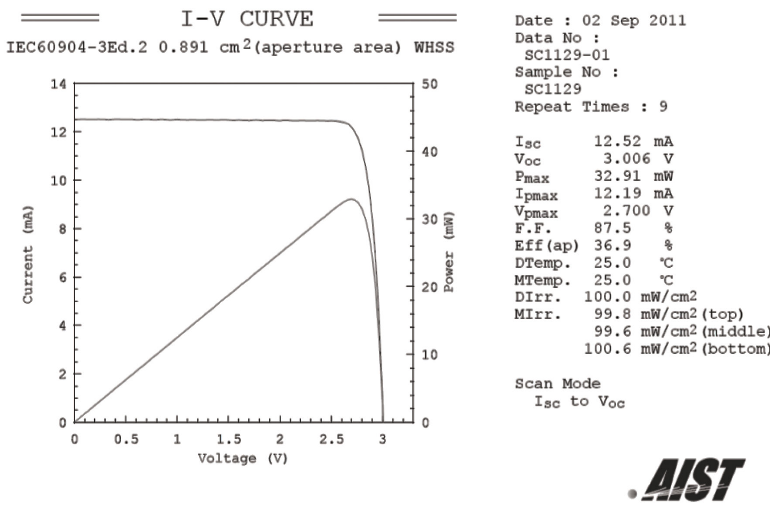


Figure 25. Current(I)–voltage(V) curve of world-record efficiency InGaP/GaAs/InGaAs measured by the AIST



Most recently, 42.1% efficiency under 230-suns has been obtained with InGaP/GaAs/InGaAs cell as shown in Figure 27 (Takamoto, et al., 2010).

FUTURE DIRECTION

Multi-junction solar cells will be widely used in space because of their high conversion efficiency and better radiation-resistance. In order to apply super high-efficiency cells widely, it is necessary to improve their conversion efficiency and reduce their cost. The new Concentrator PV system with two times more annual power generation than the conventional crystalline silicon flat-plate system will open a new market for apartment or building rooftop applications. Another interesting appli-

Figure 26. Chronological improvements in conversion efficiencies of III-V compound multi-junction solar cells under 1-sun and concentrator conditions

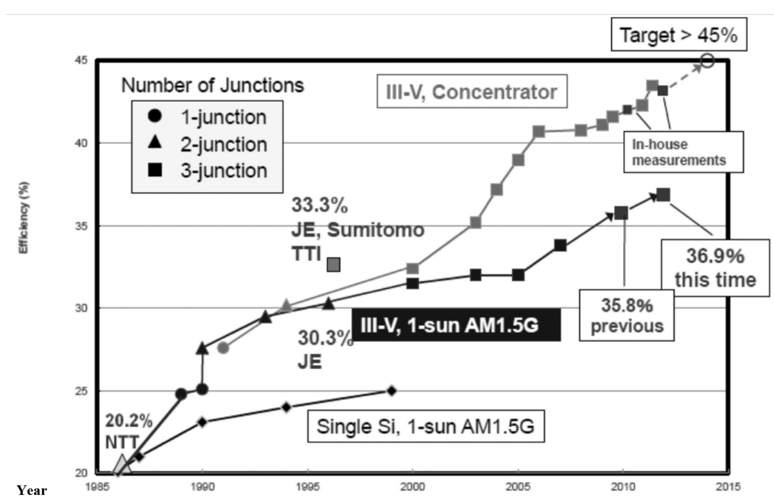
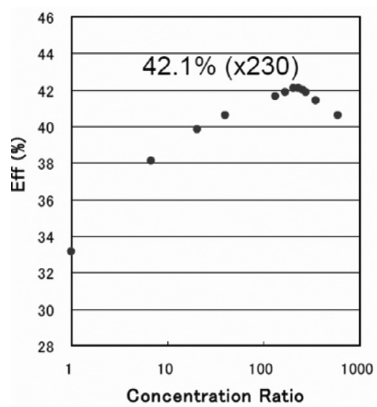


Figure 27. Concentration ration dependence of efficiency of a high-efficiency InGaP/GaAs/InGaAs 3-junction cell



cation is what we call the tree planting PV and large-scale PV power plant applications

Now, we are approaching 40% efficiency by developing concentrator MJ solar cells as shown in Figure 28. Concentrator 4-junction or 5-junction solar cells have great potential for realizing super high-efficiency of over 50% [29,30]. Therefore, concentrator 3-junction and 4-junction solar cells have great potential for realizing super high-efficiency of over 40%. As a 3-junction combination, InGaP/InGaAs/Ge cell on a Ge substrate will be widely used because this system has already been developed. The 4-junction combination of an Eg=2.0eV top cell, a GaAs second-layer cell, a material third-layer cell with an Eg of 1.05eV, and a Ge bottom cell is lattice-matched to Ge substrates and has a theoretical efficiency of about 42% under 1-sun AM0. This system has a potential efficiency of over 47% under 500-suns AM1.5 condition.

The new Concentrator PV (CPV) system with two times more annual power generation than the conventional crystalline silicon flat-plate system will open a new market for apartment or building rooftop applications. Another interesting application is what we call the tree planting PV and large-scale PV power plant applications. Since concentrator MJ and crystalline Si solar cells are expected to contribute to electricity cost reduction for widespread PV applications as shown in Figure 29, we would like to contribute to commercialization of CPV technologies as the 3rd PV technologies in addition to the first crystalline Si PV and the 2nd thin-film PV technologies. The scale of the CPV industry lags that of flat-plate PV by about one or two decades. It is expected, however, to make up this delay to the point where, in, 2013 the cumulative installed capacity will lie in the region of several hundred MWp. R&D work has to be undertaken particularly in the area of large-scale production, i.e. high-throughput, to realise this ambition. Material consumption must

Figure 28. Future predictions of solar cell efficiencies

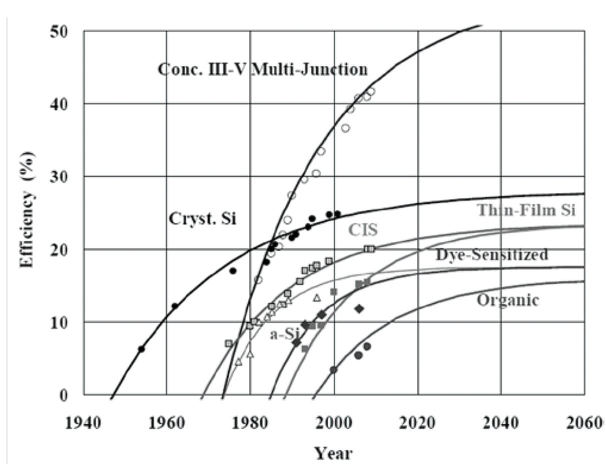
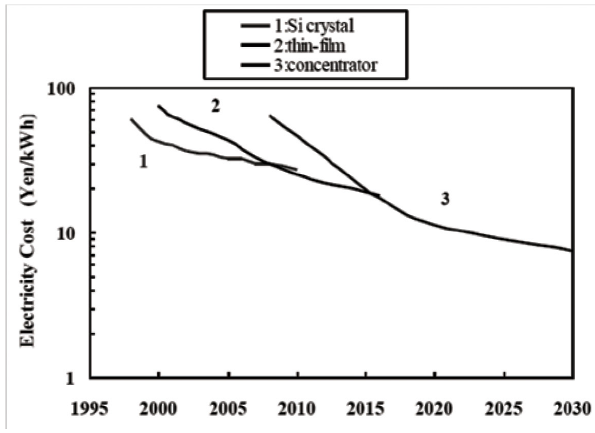


Figure 29. Scenario of electricity cost reduction by developing concentrator solar cells



also be reduced. A projection for future turnkey CPV system prices extrapolated from current prices are shown in Figure 30. Figure 30 shows the expected yearly production of Concentrator PV (CPV) systems (red) and the price of turn-key installed CPV systems in Euro/W_p (black) (EC, 2007).

The most important R&D in CPV manufacturing will aim at (EC, 2007):

1. Improving the efficiency of mass-produced cells to the levels currently seen in the laboratory (over 26%) and to 35 to 45% efficiency in the longer term
2. Improving optical elements (optical efficiency, lifetime and product engineering)
3. Automated industrial module assembly (adjustment of elements, packaging and sealing), high-throughput manufacturing with high yield, resulting in products with long lifetimes
4. Construction of light, robust, and precise trackers for all outdoor climate conditions
5. Set-up and monitoring of demonstration systems and large plants, in the range several hundred kWp (short term) to multi-MWp (medium term)
6. Techniques for guaranteeing the quality of products with intended lifetimes of over 20 years, development of standards, in-line testing, and recycling methods for the modules.

Figure 30. The expected yearly production of concentrator PV (CPV) systems (red) and the price of turn-key installed CPV systems in euro/ W_p (black)

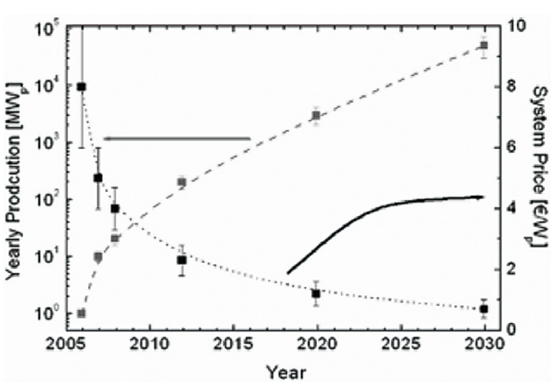
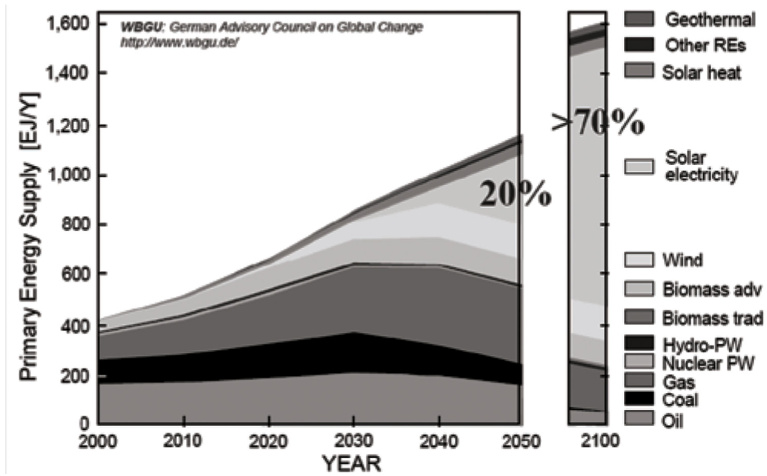


Figure 31 shows World Energy Vision 2100 recommended by the German Council on Global Change (WBGU, 2003). As a result of further development of higher efficiency, lower cost and highly reliable solar cells, larger contribution to world energy by PV is expected.

Figure 31. World energy vision 2100 recommended by the German council on global change



ACKNOWLEDGMENT

This work was supported by New Energy and Industrial Technology Development Organization (NEDO) under METI (Ministry of Economy, Trade, and Industry). The author thanks members of Toyota Tech. Inst., Sharp, Daido Steel, JAXA, JAEA, Univ. Tokyo, Meijo Univ., Kyushu Univ., and Miyazaki Univ. for their collaboration and cooperation.

REFERENCES

Ahrenkiel, R. K., Keyes, B. M., Durbin, S. M., & Gray, J. L. (1993). Paper. In *Proceedings of the 23rd IEEE Photovoltaic Specialists Conference,* (p. 42). New York, NY: IEEE Press.

Amano, C., Sugiura, H., Yamamoto, A., & Yamaguchi, M. (1987). Paper. Applied Physics Letters, (n.d), 51.

Ando, K., Amano, C., Sugiura, H., Yamaguchi, M., & Salates, A. (1987). Nonradiative e-h recombination characteristics of mid-gap electron trap in $Al_xGa_{1-x}As$ (x=0.4) grown by molecular beam epitaxy. *Japanese Journal of Applied Physics*, *26*, L266. doi:10.1143/JJAP.26.L266

Araki, K., Uozumi, H., Egami, T., Hiramatsu, M., Miyazaki, Y., & Kemmoku, Y. (2005). Development of concentrator modules with dome-shaped fresnel lenses and triplejunction concentrator cells. *Progress in Photovoltaics: Research and Applications*, *13*, 513. doi:10.1002/pip.643

Bett, A. W., et al. (2000). Paper. In *Proceedings of the 24th European Photovoltaic Solar Energy Conference*, (p. 1). WIP.

Brown, M. R., Goldhammer, L. J., Goodelle, G. S., Lortz, C. U., Perron, J. N., & Powe, J. S. … Krut, D. D. (1997). Paper. In *Proceedings of the 26th IEEE Photovoltaic Specialists Conference,* (p. 805). New York, NY: IEEE Press.

EU. (2007). *A strategic research agenda for photovoltaic solar energy technology PV technology platform*. Retrieved from http://www.eupvplatform.org/fileadmin/Documents/PVPT_SRA_Complete_070604.pdf

Fan, J. C. C., Tsaur, B.-Y., & Palm, B. J. (1982). Paper. In *Proceedings of the 16th IEEE Ptotovoltaic Specialists Conference*, (p. 692). New York, NY: IEEE Press.

Goetzberger, J., Luther, & Willeke, G. (2002). Solar cells: Past, present, future. *Solar Energy Materials and Solar Cells*, *74*, 1. doi:10.1016/S0927-0248(02)00042-9

Hutchby, J. A., Markunas, R. J., & Bedair, S. M. (1985). Paper. In *Proceedings of the SPIE, Photovoltaics*, (Vol. 543), (p. 543). SPIE.

Jackson, E. D. (1955). *Transactions of the conference on the use of solar energy 5*. Tucson, AZ: Tucson University of Arizona Press.

Khan, A., Yamaguchi, M., Bourgoin, J. C., & Takamoto, T. (2000). Room-temperature minority-carrier injection-enhanced recovery of radiation-induced defects in p-InGaP and solar cells. *Applied Physics Letters*, *76*, 2559. doi:10.1063/1.126407

King, R. R., Boca, A., Hong, W., Liu, X.-Q., Bhusari, D., & Larrabee, D. … Karam, N. H. (2009). Paper. In *Proceedings of the 24th European Photovoltaic Solar Energy Conference*, (p. 55). Munich, Germany: WIP.

Lang, D. V., Kimerling, L. C., & Leung, S. Y. (1976). Recombination enhanced annealing of the *E*1 and *E*2 defect levels in 1 MeV electron–irradiated *n* GaAs. *Journal of Applied Physics*, *47*, 3587. doi:10.1063/1.323161

Olson, J., Kurtz, S., & Kibbler, K. (1990). A 27.3% efficient $Ga_{0.5}In_{0.5}P$/GaAs tandem solar cell. *Applied Physics Letters*, *56*, 623. doi:10.1063/1.102717

Takamoto, T., Yamaguchi, M., Ikeda, E., Agui, T., Kurita, H., & Al-Jassim, M. (1999). Mechanism of Zn and Si diffusion from a highly doped tunnel junction for InGaP/GaAs tandem solar cells. *Journal of Applied Physics*, *85*, 1481. doi:10.1063/1.369278

Sugiura, H., Amano, C., Yamamoto, A., & Yamaguchi, M. (1988). Double heterostructure GaAs tunnel junction for a AlGaAs/GaAs tandem solar cell. *Japanese Journal of Applied Physics*, *27*, 269. doi:10.1143/JJAP.27.269

Takamoto, T., Agui, T., Ikeda, T., & Kurita, E. (2000). Paper. In *Proceedings of the 28th IEEE Photovoltaic Specialists Conference*, (p. 976). New York, NY: IEEE Press.

Takamoto, T., Agui, T., Kamimura, K., Kaneiwa, M., Imaizumi, M., Matsuda, S., & Yamaguchi, M. (2003). Paper. In *Proceedings of the 3rd World Conference on Photovoltaic Energy Conversion, WCPEC-3*, (p. 581).

Takamoto, T., Agui, T., Yoshida, A., Nakaido, K., Juso, H., & Sasaki, K. … Takahashi, M. (2010). Paper. In *Proceeding of 35th IEEE Photovoltaic Specialists Conference*, (p. 412). New York, NY: IEEE Press.

Takamoto, T., Ikeda, E., Agui, T., Kurita, E., Tanabe, T., & Tanaka, S. … Yamaguchi, M. (1997). Paper. In *Proceedings of the 26th IEEE Photovoltaic Specialists Conference*, (p. 1031). New York, NY: IEEE Press.

Takamoto, T., Ikeda, E., Kurita, H., Ohmori, M., & Yamaguchi, M. (1997). Two-terminal monolithic $In_{0.5}Ga_{0.5}P/GaAs$ tandem solar cells with a high conversion efficiency of over 30%. *Japanese Journal of Applied Physics*, *36*, 6215. doi:10.1143/JJAP.36.6215

Takamoto, T., Kaneiwa, M., Imaizumi, M., & Yamaguchi, M. (2005). InGaP/GaAs-based multijunction solar cells. *Progress in Photovoltaics: Research and Applications*, *13*, 495. doi:10.1002/pip.642

Tomita, T. (2006). Paper. In *Proceedings of the 4th World Conference on Photovoltaic Energy Conversion*, (p. 2450). New York, NY: IEEE Press.

WBGU. (2003). *World in transition – Towards sustainable energy systems*. Retrieved from http://www.wbgu.de/

Wolf, M. (1960). Detailed balance limit of efficient p-n junction solar cells. In *Proceedings of IRE 48*, (pp. 1246). IRE.

Yamaguchi, M. (2003). III–V compound multi-junction solar cells: Present and future. *Solar Energy Materials and Solar Cells*, *75*, 261. doi:10.1016/S0927-0248(02)00168-X

Yamaguchi, M. (2004). Paper. In *Proceedings of the 19th European Photovoltaic Solar Energy Conference*. Munich, Germany: WIP.

Yamaguchi, M., & Amano, C. (1985). Numerical analysis for high efficiency GaAs solar cells fabricated on Si substrates. *Journal of Applied Physics*, *58*, 3601. doi:10.1063/1.335737

Yamaguchi, M., Okuda, T., Taylor, S. J., Takamoto, T., Ikeda, W., & Kurita, H. (1997). Superior radiation-resistant properties of InGaP/GaAs tandem solar cells. *Applied Physics Letters*, *70*, 1566. doi:10.1063/1.118618

Yang, M.-J., Yamaguchi, M., Takamoto, T., Ikeda, E., Kurita, H., & Ohmori, M. (1997). III–V compound multi-junction solar cells: Present and future. *Solar Energy Materials and Solar Cells*, *45*, 331. doi:10.1016/S0927-0248(96)00079-7

This work was previously published in Advanced Solar Cell Materials, Technology, Modeling, and Simulation, edited by Laurentiu Fara and Masafumi Yamaguchi, pp. 139-162, copyright 2013 by Engineering Science Reference (an imprint of IGI Global).

Chapter 109
Zero-Downtime Reconfiguration of Distributed Control Logic in Industrial Automation and Control

Thomas Strasser
AIT Austrian Institute of Technology, Austria

Alois Zoitl
Vienna University of Technology, Austria

Martijn Rooker
PROFACTOR GmbH, Austria

ABSTRACT

Future manufacturing is envisioned to be highly flexible and adaptable. New technologies for efficient engineering of reconfigurable systems and their adaptations are preconditions for this vision. Without such solutions, engineering adaptations of Industrial Process Measurement and Control Systems (IPMCS) will exceed the costs of engineered systems by far and the reuse of equipment will become inefficient. Especially the reconfiguration of control applications is not sufficiently solved by state-of-the-art technology. This chapter gives an overview of the use of reconfiguration applications for zero-downtime system reconfiguration of control applications on basis of the standard IEC 61499 which provides a reference model for distributed and reconfigurable control systems. A new approach for the reconfiguration of IEC 61499 based control application and the corresponding modeling is discussed. This new method significantly increases engineering efficiency and reuse in component-based IPMCS.

INTRODUCTION

The decisive factor for the market success of the manufacturing industry (e.g. auto manufacturers and part makers, process industry etc.) is a fast and flexible reaction to changing customer demands—companies must show a high degree of changeability. New paradigms like "Flexible production up to small lot-sizes", "Mass Customization" or "Zero-Downtime Production" will achieve these requirements but demand completely new technologies for its realization (European Commission, 2004). Changeability, which describes the ability

DOI: 10.4018/978-1-4666-1945-6.ch109

of companies being flexible concerning customer demands, impacts all levels of product manufacturing. In particular these are the agility at a strategic level, the transformability at a factory level and the reconfigurability at the manufacturing system and machine level (Koren, 1999).

The state-of-the-art in manufacturing systems is inadequate to meet the above mentioned requirements. Current manufacturing systems are either tailored towards a specific product at high volume production and thus they can hardly be adapted to new products or they are flexible and programmable but technology specific and only for single item or small batch production. Another relatively new approach which is flexible and programmable but less technology specific but also hardly adoptable concerning the above mentioned requirements is the usage of “Multi Machining Technology Integration Production Systems” (Abele, 2005) which are characterized by the static implementation and combination of different technologies within one production system. The major drawback of this approach is that it is very resource consuming and therefore it can hardly be ported to small and resource constrained embedded controllers which are often used in modern industrial automation and control systems.

To reach the above mentioned changeability at the manufacturing systems and machine level it can be postulated that a change from product and technology rigid manufacturing systems towards product and technology flexible, modular, easy to setup component-based production systems is necessary. Following consequently this trend means that future plants will produce their products on manufacturing systems and machines which will be designed and setup just prior to production of goods since they are constructed on basic building blocks. Such building blocks are in general smart mechatronic components with embedded intelligence. These building blocks are designed in a way that they provide a specific manufacturing and/or automation functionality and they are not reconfigurable in general. Machining, assembly and transport systems of such production systems are also designed and set up by the utilization of various flexible autonomous and intelligent mechatronic components just before usage within the production line.

The consequences of the above mentioned attempt are extensive and many technological breakthroughs will be necessary. Beside others the development of an adequate automation system for heavily interacting distributed real-time systems can be seen as a major task. Current architectures of IPMCS do not conceptually support reconfiguration and distribution which are necessary to fulfill the requirements for the above mentioned systems (Sünder, 2006). Distributed embedded real–time systems for industrial automation and control of plants that evolve towards zero-downtime adaptable systems will play a key role to realize the roadmaps towards adaptive manufacturing (Koren, 1999) of products, goods and services in 2020. Most value will then be added in engineering and performing a system transition or reconfiguration (the change from one system state to another) rather than in engineering and performing “normal operation”.

The challenge and aim of this chapter is to present an approach for modeling of reconfiguration control applications based on the IEC 61499 standard which provide an adequate engineering methodology for programmed system reconfiguration (i.e. system reconfiguration executed by a special control application). The first section discusses general reconfiguration issues. The next section gives an overview of state-of-the-art in reconfiguration of control applications. After that a summary of the main features and characteristics of IEC 61499 as the reference model for distributed automation and control systems with special focus of the management capabilities for reconfiguration is given. Furthermore an enhanced IEC 61499 Device Management is introduced which is used for the proposed approach for the controlled reconfiguration of control applications as the main topic of this chapter. Especially the modeling of reconfiguration control applications based on the IEC 61499 reference model is topic of this chapter. Tests at the Odo Struger Laboratory at Vienna University of Technology (Automation and Control Institute) and at PROFACTOR are also presented. The summary of this work concludes this chapter.

BACKGROUND: GENERAL CONSIDERATIONS ABOUT RECONFIGURATION

In industrial informatics, reconfiguration of software modules and components has been discussed many times, for instance (Brennan, 2002; Bussmann, 1999; Burmester, 2004). Agile software processes like the extreme programming method recognize change as an essential part of the software life–cycle (Mens, 2005).

Basic vs. Dynamic Reconfiguration

Within the scope of this chapter reconfiguration is described best as altering a system's operation in order to meet changes in requirements. Quite a simple reconfiguration method—hence further denoted as basic reconfiguration—is to stop current operation, apply all necessary changes to the system and restart the desired operation again (coldstart). But this approach is not sufficient to meet the constraints mentioned in the introduction.

Brinkschulte et al. (2005) define dynamic reconfiguration as reconfiguration of an application while it is running. What therefore further distinguishes dynamic from basic reconfiguration is provision for timeliness constraints (Brennan, 2002). Dynamic reconfiguration considers timeliness as a crucial facet of correctness. With respect to control applications this means that timeliness requirements put on the whole application must be preserved while parts of it are modified.

Classification of Dynamic Reconfiguration

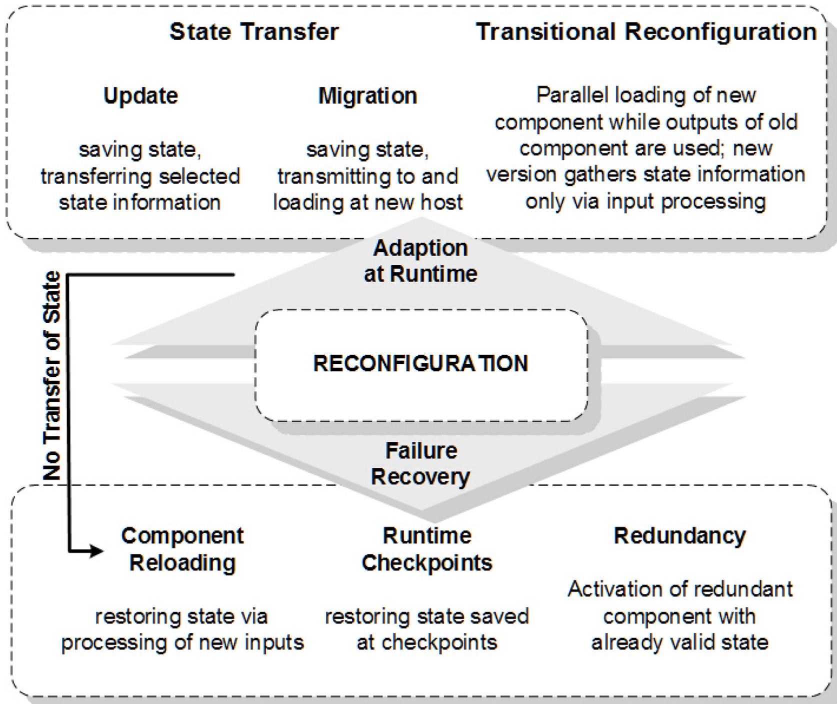
Within the topic of dynamic reconfiguration a big variety of opportunities arises. This especially demands means of classification for the different approaches within this field. Wermelinger (1997) has formulated 3 essential questions that have to be answered for dynamic reconfiguration:

1) What kind of modifications can be done?
2) How are they performed?
3) When may the changes be performed?

The first two questions concern the reconfiguration model (model of system architecture and modification process) as well as the features of the underlying operating system and/or the runtime environment. Question 3 is of special interest. The answer to this question is that those parts of an automation system that will undergo changes have to be in a consistent state before. But consistency is not only is a prerequisite for reconfiguration. Quite on the contrary, consistency is an important attribute of systems that must be preserved over reconfiguration in order to avoid unnecessary disruption to the system environment. This requires transfer of state information from the current system state to the desired state after reconfiguration.

Rasche and Polze (2005) give an idea of how consistent state information can be ensured in the context of different use cases of reconfiguration, illustrated in Figure 1. Adaptation at runtime (upper part of the figure) either requires transfer (save/transmit/load) of state information (update, migration) or gathering state information via input processing (transitional reconfiguration). On the other hand failure recovery (lower part of figure) must be able to restore a consistent state either by loading saved state information (runtime checkpoints) or input processing (component reloading) or by activating redundant components (which already have gathered valid state information).

Figure 1. State consistency and reconfiguration



State-of-the-Art Reconfiguration Approaches

This section briefly outlines current reconfiguration approaches relevant for industrial automation. Today common automation systems predominantly rely on the programming paradigms of the standard IEC 61131–3 (IEC, 2003) dedicated to systems based on programmable logic controllers (PLCs). The common paradigm within IEC 61131 is a device centered engineering methodology. Aspects of distribution (cooperation of several devices within an overall application) have to be considered from the beginning of the engineering phase.

Current IEC 61131–3 based engineering tools such as (kirchner SOFT, 2009) already enable online code exchange—called instant-reload—for single devices at task level, including transfer of variable values. Their product logi,CAD already provides some basic support for reconfiguration of applications (i.e. of programs, resources and configurations; this keywords are used in terms of IEC 61131) in the programming system logi.CAD as well as in kirchner SOFT's target system implementations (i.e. Soft-PLCs). Their reconfiguration approach is centered on managing changes for a single IEC 61131-3 resource at a time. Changing applications at runtime without stopping them is fully supported for the user. Unfortunately using such a resource-centered approach has some disadvantages when changes for too many different applications running in different resources have to be applied simultaneously. In addition there exist also some other drawbacks of this approach like undeterministic switching points in time (due to cyclic execution policy), lack of fine granularity (reconfiguration at task level), jitter effects (task reconfiguration influences other tasks) or the possibility of inconsistent states (e.g. leading to deadlocks) (Sünder, 2006).

The International Electrotechnical Commission (IEC) addresses these issues by the standard IEC 61499 (IEC, 2005), which extend the Function

Block (FB) model of IEC 61131–3 in order to meet the requirements of engineering distributed automation systems. The most important concepts of IEC 61499 are an event–driven execution model; a management interface capable of basic reconfiguration support and the application centered engineering methodology. The major benefit of this approach is a separation of concerns: first the whole application is programmed as a function block network like in centralized systems; afterwards the components of this network (the function blocks) are mapped to the devices of the real system where they shall be executed. Doing so e.g. facilitates the movement of functionality from one controller to another as mentioned in section I, since only the mapping is concerned while the original application remains unchanged. A more detailed description of IEC 61499 and its aptitude for reconfiguration purposes can be found in (Strasser, 2005).

Various researchers have developed reconfiguration methodologies based on IEC 61499: the research project TORERO (Lorentz, 2006) focuses on plug–and–play and self–(re)configuration of field devices in a so–called total lifecycle approach utilizing IEC 61499 for modeling control logic. The main objective of this project was on the creation of a total lifecycle web-integrated control design architecture and methodology for distributed control systems for industrial and factory automation. The major focus and research was on the development of a self-configuring and maintaining control infrastructure with automated distribution support. One of the major drawbacks of the TORERO approach is the fact that a device hast to be stopped for the deployment of code and afterwards it has to be restarted.

Thramboulidis et al. (2004) use Realtime–UML (Unified Modeling Language) as a meta model between the design models of IEC 61499 and their implementation models to support dynamic reconfiguration of control applications. Brennan et al. (2002) propose an agent–based reconfiguration approach, extending the IEC 61499 function block model: they introduce a second state–machine within the basic function block together with a reconfiguration agent and a second event and data flow for reconfiguration purposes. All these reconfiguration approaches rely on technology that is outside the scope of IEC 61499 (e.g. Universal Plug and Play, agents or Realtime–UML), they do not consider utilizing the basic reconfiguration functionality provided by the IEC 61499 management model outlined in the next section.

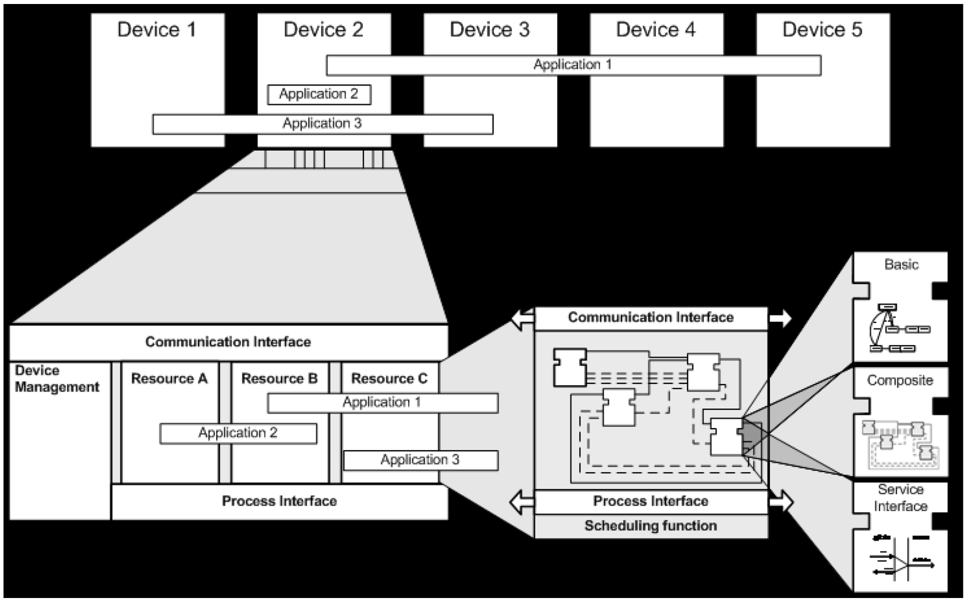
IEC 61499 as Reference Model for Reconfigurable Distributed Control

The new standard IEC 61499 (IEC, 2005) serves as a reference architecture that has been developed for distributed, modular, and flexible control systems. It specifies an architectural model for distributed applications in IPMCS in a very generic way and extends the function block model of its predecessor IEC 61131-3 (Function Block Diagram - FBD) with additional event handling mechanisms and concepts for distributed systems.

The standard builds a good basis to overcome the above mentioned problems according to reconfiguration processes in current IPMCS. The following sections describe some fundamental issues of IEC 61499 that make this standard suitable as a reference architecture for building zero-downtime IPMCS by using the concept of reconfiguration applications. More detailed information is available from (Christensen, 2009; Lastra, 2004; Lewis, 2001).

Figure 2 provides a brief overview of the most important concepts and models of the IEC 61499 standard. In the following we briefly explain important aspects of the reference model in order to better understand this standard. For more details about definitions, concepts and models we refer to the IEC standard (IEC, 2005).

Figure 2. The IEC 61499 reference model for distributed automation and control



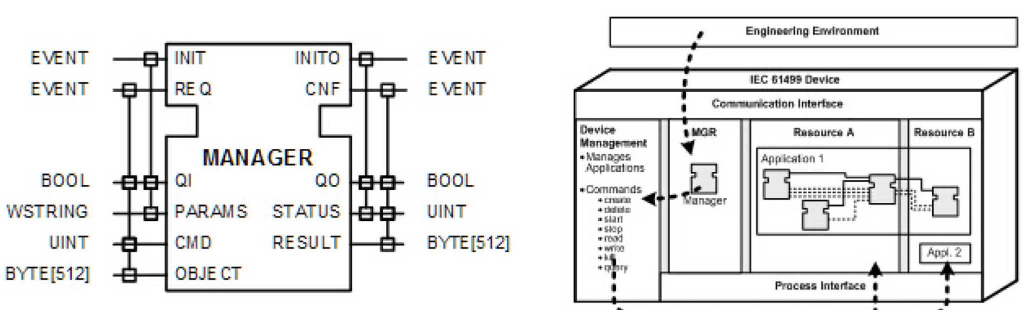
Execution Control by Events

The function block model of IEC 61499 defines function blocks that are characterized by the occurrence of two Input/Output (I/O) types: events and data. The execution is triggered by events, needed data have to be valid at the inputs of the FB before an event arrives (denoted as the WITH construct in IEC 61499, i.e. the vertical lines between event and data inputs or outputs as shown in Figure 3a). Only if an event occurs at a FB input, the runtime environment has to process the execution of this FB. Therefore it is obvious and clearly understandable whether an application is active or not and even exactly which part is currently processed. Compared to PLC-based approaches like IEC 61131-3 the IEC 61499 approach doesn't define any time base nor has a cyclic-based execution approach

Distribution Model

One of the main intentions of IEC 61499 is enabling distributed applications. The system model

Figure 3. Generic management function block type



(a) (b)

consists of devices connected by a communication network and applications. These applications may be distributed within devices. Moreover the IEC 61499 standard allows in addition the separation of devices into independent resources which are execution containers for function block networks. This concept allows a local distribution of an application part of a specific device into several resources of this device. The engineering process starts with the top-level functionality that has to be realized without reference to the concrete hardware structure. By use of libraries of software components, the user models the needed applications. The last step within the engineering cycle is the mapping of the applications to concrete hardware components; independent of whether the application is executed by one device or distributed to several devices.

Management Interface of IEC 61499 Devices

The configuration of a distributed automation and control system based on IEC 61499 can be enabled by the use of management functions which can be included in each device. For this purpose the standard defines a management application, represented by a management FB (the generic interface is depicted in Figure 3a). By using this FB combined with a remote application, mutual access between different IEC 61499 compliant devices is possible. The IEC 61499 Compliance Profile for Feasibility Demonstrations (available from Christensen (2009)) describes a concrete interface of a management FB and an appropriate remote application. The following standardized management functions can be used to interact with a device (IEC, 2005, Part 1-Table 6, 8). For illustration examples of possible actions are added.

- **CREATE:** FB instance, data or event connection
- **DELETE:** FB instance, data or event connection
- **START:** FB instance, application
- **STOP:** FB instance, application
- **READ:** FB instance data outputs
- **WRITE:** FB instance data inputs
- **KILL:** FB instance
- **QUERY:** FB types, FB instance, data or event connection

Especially the management of software components (i.e. FBs) regarding to their execution is a very important issue in reconfiguration processes. A FB instance operates according to a state machine (IEC 2005, Part 1-Figure 24) that includes an initialization and operation state controlled by management commands: Create, Start, Stop/Kill and Delete.

Verification and Validation

A very important feature of a control modeling language is the possibility to proof correctness. The basic function block of IEC 61499 includes an event-triggered state machine (Execution Control Chart, ECC) that provides the connection between an incoming event and the execution of algorithms and outgoing events. This basic model enables the conversion of IEC 61499 applications to mathematical models like Petri Nets to provide verification and validation. Different approaches to this topic are already available, for instance Net Condition/Event Systems (Vyatkin, 1999), timed automata (Stanica, 2004) or Signal/Net Systems (Schnakenbourg, 2002).

ENHANCED IEC 61499 DEVICE MANAGEMENT FOR DOWNTIMELESS RECONFIGURATION

In the case of dynamic reconfiguration of control applications or application parts it is necessary to define several (re-)configuration services, which provide enough functionality to vary FB parameters, to change the behavior of FBs or Function Block Networks (FBN) and to create or delete instances of FBs and Resources. This section describes the necessity of an enhanced set of (re-) configuration services based on the IEC 61499 standard. As mentioned in the previous section the IEC 61499 standard already provides a basis set of (re-)configuration services but it lacks a method and an implementation how these services can be efficiently applied for the reconfiguration of control applications. Therefore the aim of this approach is to make extensions to the IEC 61499 standard to simplify the reconfiguration process. This will be explained in the next sub-sections.

Considerations on Enhanced (Re-)Configuration Services

The following list displays possible reconfiguration scenarios for IEC 61499 based FBNs:

- **Instance reconfiguration:** regards to the exchange, adding or removal of a certain instance of IEC 61499 types (FB, Data, Resource, and Device). In case of an exchange of FB instances there exists the possibility with and without the adoption of internal FB variable values.
- **Type reconfiguration / runtime set-up:** the supported IEC 61499 types (Function Block, Data, Resource and Device) have to be exchanged; added or removed from a runtime system (i.e. this issue is mainly related to the change in the structure and elements of the type library of the runtime system).
- **Parameter reconfiguration:** the parameterization of Function Blocks, Resource or Device instances has to be adapted.
- **Application structure reconfiguration:** the logical configuration of the control applications has to be adapted. This is mainly related to the change of event and data connections or the type instances in IEC 61499 based control applications.
- **System reconfiguration:** the structure of the IEC 61499 system (i.e. the device and communication structure) is adapted to new requirements. This issue means that the IEC 61499 system configuration (i.e. the system model as represented in Figure 2) will be changed. Such changes can be the addition or deletion of devices and/or resources or even the exchange of existing devices with other once. Moreover it is also possible to (ex-)change the communication system similar to the (ex-)change of the device configuration.

Categorization of Basic Reconfiguration Services

Based on the device management commands introduced by IEC 61499 and the analysis of general reconfiguration tasks the following set of basic reconfiguration services are identified and categorized:

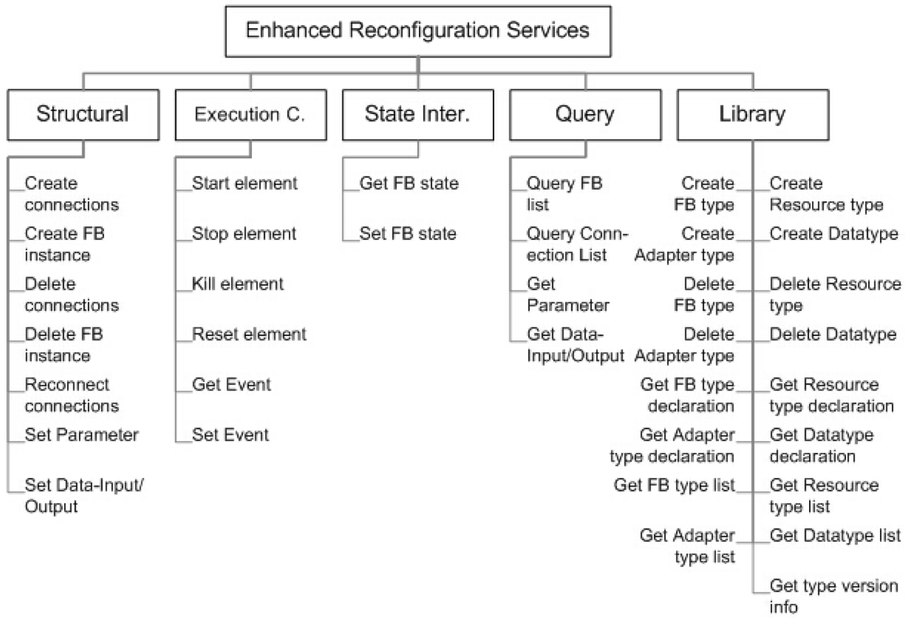
- **Structural services:** These services are mainly related to the modification/change of the logical configuration/structure of control applications and their behavior. IEC 61499 already provides a very rudimentary set of services that allow to create and delete FBs, resources and connections and to write parameters of FBs and resources. This rudimentary set of IEC 61499 configuration services is extended with additional features to form a more comprehensive set of basic structural reconfiguration services.
- **Execution control services:** These services allow interaction with the currently executing control application. In this case IEC 61499 provides a start, stop and kill mechanism for FBs. By extending these mechanisms with an additional restart ser-

vice and also using these services for event and data connections the reconfiguration process provides more possibilities to manage and change the execution of Function Block Networks (i.e. IEC 61499 applications). This especially enables a more efficient rewiring of FBs without use of extra computational power compared to the standard execution control services (i.e. start, stop and kill) provides by the IEC 61499 standard. The execution of only one management service (i.e. restart) is necessary compared to the execution of the service sequence stop FB and start FB. The restart service can be implemented more efficiently in a runtime environment. Fulfilling this requirement is essential if running control applications are reconfigured. In the case of reconfiguration of running control applications high effort has to be put into keeping the interruption of the running system as short as possible.

- **State interaction services:** These services retrieve and set internal states of FBs. The IEC 61499 standard does not clearly define these services. Such services are mainly needed for providing transition management, which takes care that the process is also under stable control during the reconfiguration phase (i.e. instabilities during the reconfiguration phase have to be avoided). The FBs are just components without access to anything outside their interface. In the case of reconfiguration full access to internal FB states and data is necessary to force the FB to a defined state avoiding reinitializing the FB.
- **Query services:** These services are necessary to acquire the current state of the distributed control application running on several devices. Especially during dynamic reconfiguration the system state will change several times. Because of that, query services are necessary in order to obtain the actual systems state and moreover also consistency checks are necessary to during dynamic reconfiguration. A consistent state after a reconfiguration is very important in industrial automation and control systems because inconsistent system states can lead to catastrophic faults. Another fact is that the execution of reconfiguration applications is characterized by a time span between engineering and execution. Within a multi user system, the reconfiguration has to ensure that the considered system part has not changed between this time span. The query mechanisms provided by IEC 61499 can serve as basis for this service.
- **Library services:** These services allow adding or removing IEC 61499 types such as Function Blocks, Event, Data or Resources to or from the runtime library of the control device respectively. They are also not clearly defined within IEC 61499. The main description of FBs is defined by the use of the Extended Mark-up Language (XML), but e.g. the description language for algorithms within FBs is still not defined. The capability of device independent type libraries needs a full new specification.

In addition to the above mentioned basic reconfiguration services it is often necessary to define further services that will be used in many reconfiguration applications (e.g. rewire a connection). To simplify reconfiguration applications and to reduce their size, it is envisaged to provide also more sophisticated reconfiguration services in addition to the basic services whereby a trade-off between reconfiguration application size and runtime environment complexity has to be found. Figure 4 gives an overview of the set of identified basic reconfiguration services based on the above categorization.

Figure 4. Enhanced Reconfiguration Service of IEC61499 Device Management



The above-presented enhanced services are either implemented as IEC 61499 XML commands or as special management FBs.

Reconfiguration of Control Applications

The reconfiguration can be initiated from a user via configuration tools (see Figure 5) or as a special control application—herein referred as Reconfiguration Control Application (RCA)—as depicted in Figure 6.

Therefore in both cases an underlying runtime system has to provide the necessary (enhanced) services for reconfiguration which are presented above. The reconfiguration steps can also be directly implemented into the control application but the RCA is especially used to encapsulate

Figure 5. Reconfiguration via configuration tool

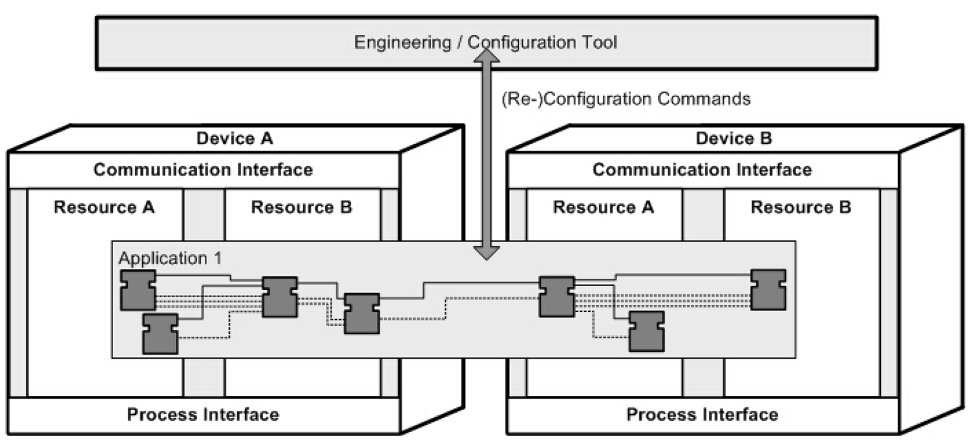
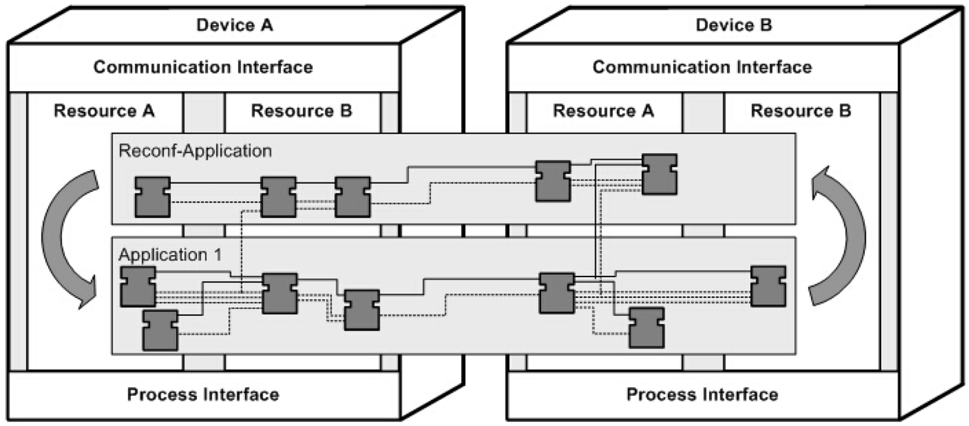


Figure 6. Reconfiguration Control Application



these reconfiguration steps into a separate application. This makes the presented approach for a control engineer more useable otherwise the control application will be overloaded with reconfiguration parts. The RCA is built of special IEC 61499 compliant management FB types that encapsulate the above mention reconfiguration services for dynamic reconfiguration of control application in a distributed system. The trigger to execute the RCA can either be initiated by the user, the control application or by the underlying controlled process.

AN ENGINEERING METHOD FOR RECONFIGURATION OF IPMCS APPLICATIONS

To overcome the limitations of current embedded industrial automation and control engineering methods, we propose an application centered engineering method for efficient component-based modeling of applications for controlled reconfiguration of IPMCS. The execution of a reconfiguration will become a normal operational state of such systems, while dependability and Quality of Service (QoS) of the reconfigured system are not endangered and migration is cost–efficient. Thus important steps towards Adaptive Manufacturing will be realized.

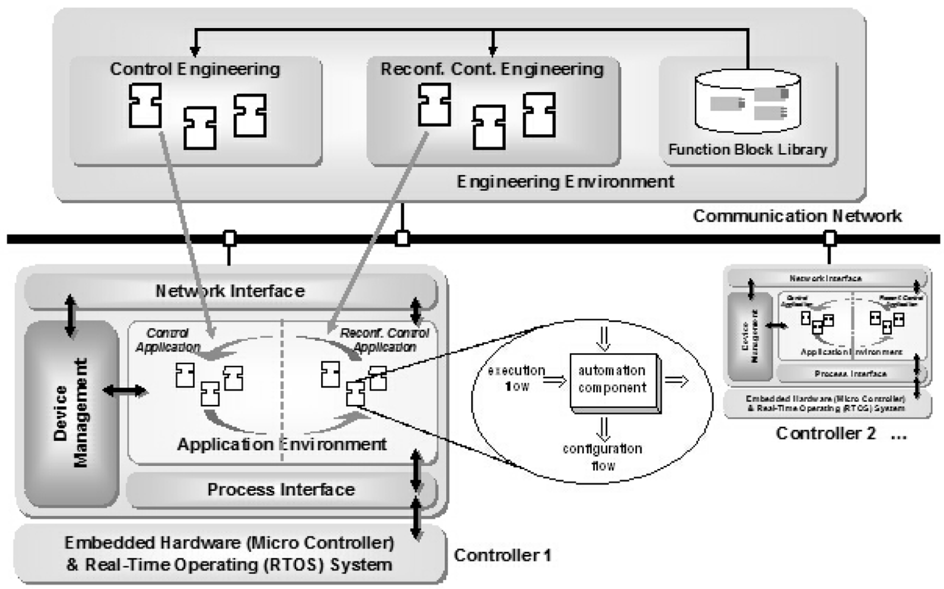
An Approach for Controlled Reconfiguration

The top–level approach focuses on replacing state–of–the–art “ramp down—stop—download—restart—ramp up” methods with a zero-downtime system reconfiguration, which is controlled by a reconfiguration control application that is modeled with components in the same way as control applications are. Zero-downtime means in this context to perform the controlled reconfiguration during runtime without the need to stop the execution of a runtime system. Reconfiguration control can be either executed from an engineering environment or it can be distributed to different controllers (as shown in Figure 7).

Modeling Cycle for Control and Reconfiguration Control Applications

The following engineering process provides a method for handling system reconfiguration of IPMCS by an efficient support for engineering of control applications and reconfiguration control applications. This engineering method consists of the following four major parts as depicted in Figure 8.

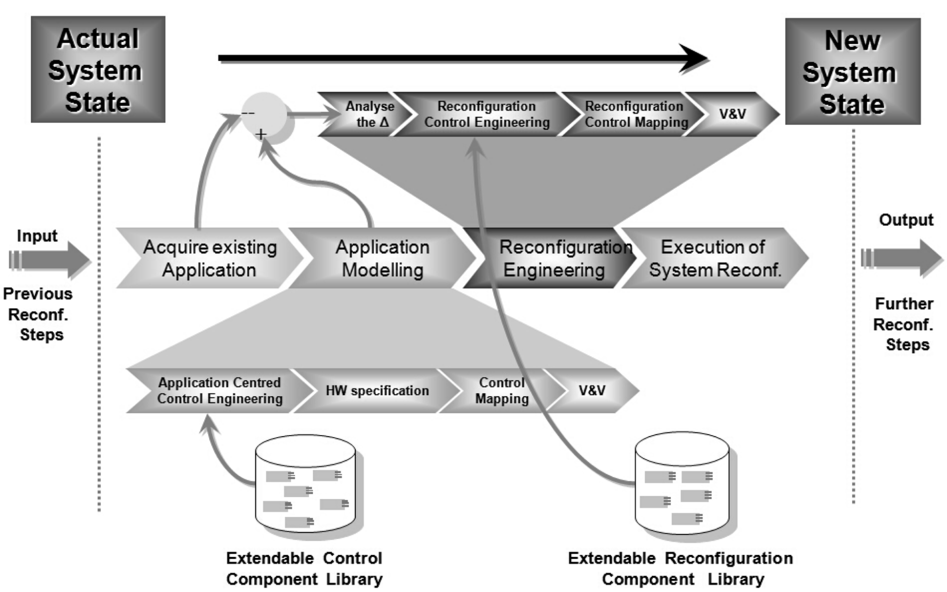
Figure 7. Controlled reconfiguration of control applications



1. Acquire Current System State (Acquire existing application)

Collect all data necessary for describing the current system state and deliver it as input to the application modeling. The data consists of the system model including applications currently running in the system, the hardware configuration of the system (used devices and network structure), the mapping of the applications to the different devices and the hardware capability descriptions.

Figure 8. Modeling cycle for control and reconfiguration control applications



2. Model New Control Application (Application Modeling)

Modeling of the new control application based on the existing ones by adding/removing components, their interconnections and specification of application properties (e.g. real–time constraints). The next step is the configuration of the system configuration (i.e. the IEC 61499 system model) with Devices, Resources and Network connections. After this the modeled application parts are mapped to the according devices they should be executed on. The final step is a formal verification and validation of the control application n order to determine if the specified application constraints will be met or not. For this verification and validation step approaches like Net Condition/Event Systems (Vyatkin, 1999), timed automata (Stanica, 2004) or Signal/Net Systems (Schnakenbourg, 2002) can be applied.

3. Model Reconfiguration Control Application (Reconfiguration Engineering)

Differences between the current running control application and the newly modeled control application are determined with the Delta Analysis. These differences serve as input and starting point for the modeling of the reconfiguration control application that will change the existing application to the new one. The reconfiguration control properties and parameters are specified in the same way as control application properties (according to step 2). Similar to the mapping of the control application the reconfiguration application parts are mapped to the devices. The final step is the verification of the reconfiguration control application together with device capabilities in order to determine if reconfiguration constraints (e.g. the timeframe in which a reconfiguration can take place or how long a reconfiguration can take place, etc.) will be met and the running application is disturbed as little as possible.

4. Execute Reconfiguration Control Application (Execution of System Reconfiguration)

To execute reconfiguration control applications the utilization of basic reconfiguration services at runtime platforms based on IEC 61499 management commands is necessary. The first step is to instantiate the reconfiguration control application on the according devices. Next the execution of the reconfiguration control application is done, i.e. the currently running control application is transformed into the new control application. This is achieved through basic reconfiguration services proving real–time constraint execution of reconfiguration processes at device level. The main services are:

- Control application component load/remove
- Connect/disconnect
- Query parameters
- Write parameters
- Query component state
- Write component state

To finish the reconfiguration procedure the reconfiguration control application will be removed after it has successfully executed all commands and all changes are finished.

Reconfiguration Methodology

The process of reconfiguring an automation system without downtimes sets high demands on the underlying concepts and methodologies: Applications within the automation system have to work without disturbances; the reconfiguration process has to be adapted to the special environmental conditions of the affected application part. Any fault during the reconfiguration process has to be managed at least to such a degree, that the system is left in a defined state. As described in the introduction the standard IEC 61499 already includes basic management commands for configuration and reconfiguration of applications. This section now proposes useful extensions to the management model of IEC 61499 enhancing the reconfiguration abilities of IEC 61499 with special regard to remaining compliant to the ideas of the standard.

The main idea of this new methodology is to control the reconfiguration of control logic (part of the control application) by an application, the so called reconfiguration application. The reconfiguration application does not differ from any other application and can therefore be modeled in the same manner. But it will have an impact on the control application, which means that it will include management commands that influence FBs and their interconnections in a defined manner. The basic idea of reconfiguration applications is described by Sünder et al. (2006). The presented engineering methodology builds on these results, particularly with regard to the setup and characteristics of reconfiguration applications.

Reconfiguration Control Terminology and Definitions

At this point it is suitable to define a few terms used later on for the development of an appropriate engineering methodology for dynamic reconfiguration processes:

- **Region of Interest (ROI):** this is a specific part of an IEC 61499 application that will be target for a specific reconfiguration. A reconfiguration application is assigned to this specific part. The definition of this region is given by the following cases of system changes:
- **Creating/Deleting a function block:** the suggested ROI is the function block itself plus the connected event as well as data inputs and outputs, since there is a significant temporal order of operations (a function block cannot be deleted unless all of its connections have been deleted).
- **Creating/Deleting a connection:** the suggested ROI consists of the corresponding halves of source and target function blocks of the connection along with their interconnections.
- **Creating/Deleting a parameter (i.e. a parameter in the context of IEC 61499 is a fixed value on a data input of a Function Block):** quite similar to connections, but without a source function block. So the suggested ROI only consists of the left half of the target FB and the parameter itself.
- **System reconfiguration:** the whole process of transforming the current control application in an IPMCS into the desired control application (instead of the state–of–the–art method "ramp down—stop—download—restart—ramp up"). Usually such a process is carried out by a number of reconfiguration applications on a number of ROIs. Reconfiguration takes place in four phases as depicted above (see Figure 8).
- **Reconfiguration Execution Control Function Blocks (RECFB):** special function blocks encapsulating reconfiguration functionality (e.g. a whole reconfiguration application). Within the execution of *RECFB* 3 reconfiguration control execution phases can be distinguished (see Figure 9; the vertical connections of the event as well as data inputs and outputs denote that only such data inputs and outputs are sampled which are connected to a specific event; this is a general concept of the IEC 61499 Function Block model; more details are explained in the IEC 61499 standard (IEC, 2005)):
- **INIT Sequence:** is needed in order to start-up and to configure the RECFB and the underlying runtime environment.
- **RINIT Sequence:** contains all preparation work that is needed to enable a controlled change of applications (for in-

stance new FBs will be instantiated and interconnected).

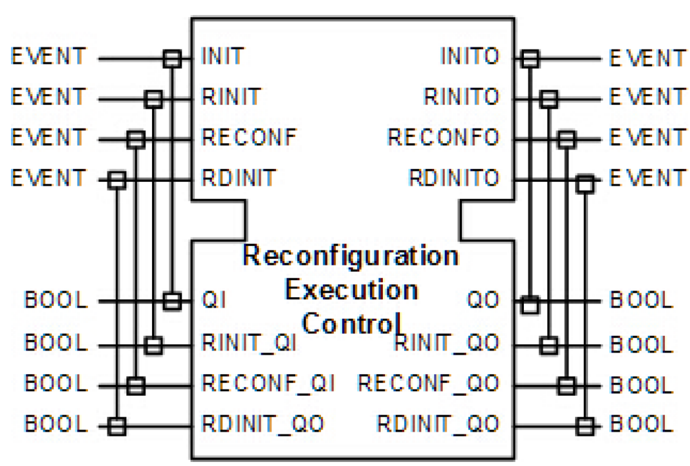
- **RECONF Sequence:** encapsulates all operations necessary to change from the old state to the new one. This for instance includes reading of internal states, calculation of the new internal states and writing the internal state to the new FB in case of a simple FB exchange with awareness to consistency. The final step of this sequence is the switch to the new FB.
- **RDINIT Sequence:** responsible for cleaning up after reconfiguration (e.g. deleting connections and function blocks not needed anymore). One *RECFB* belongs to one ROI, there is a one to one relation between them.
- **Reconfiguration Control Application:** application (i.e. a special IEC 61499 based function block network) built of *RECFBs*, that carries out the whole system reconfiguration process on an ordinary IEC 61499 system.

Reconfiguration Engineering Method

Reconfiguration applications define and describe the controlled reconfiguration within a ROI during runtime. The main intentions for the introduction of a reconfiguration engineering method handling these reconfiguration applications are:

- **Complexity:** The use of reconfiguration applications includes a reasonable amount of engineering complexity. The bigger the ROI is the more complex the whole reconfiguration application will be. This can lead to hardly understandable reconfiguration control applications which make the development and maintenance of them more complex. By using small or smaller ROIs, the engineering complexity keeps quite low.
- **Scalability:** There has to be a way to compose larger reconfiguration scenarios from several smaller reconfiguration steps. The RECFBs give the possibility to build composite RECFBs in the same way as in “normal” control applications using the same means (i.e. the models and concepts of the IEC 61499 standard).
- **Reusability:** As scaling is possible, the use of libraries for often used scenarios within typical application scopes will be useful. The concept enables such standardized RECFBs and their simple usage within composite RECFBs, again in the same way as in “normal” applications.

Figure 9. Reconfiguration Execution Control Function Block (RECFB)

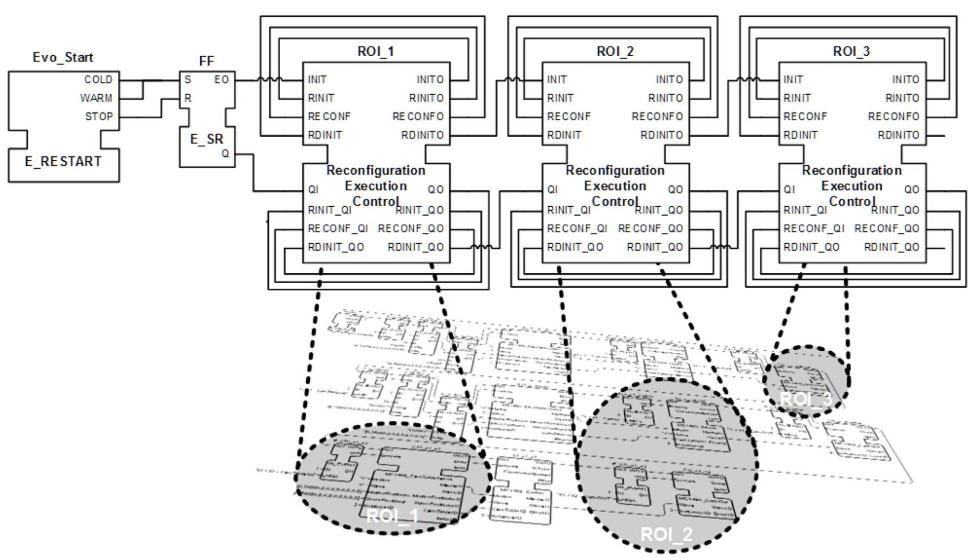


The main idea of the proposed approach is to maintain the fundamental principles of IEC 61499 (application centered engineering, component-based architecture, event–driven execution and the management model) for reconfiguration engineering. Figure 10 illustrates a reconfiguration control application containing 3 RECFBs. Each of these RECFBs contains a reconfiguration application performing reconfiguration on the associated ROI in the IEC 61499 application shown in the lower part of the figure. The interconnection of the RECFBs defines the way of executing the three steps. In the shown example only one RECFB is active at the same time (serial execution). In detail, each RECFB runs through the three phases of Initialization, Reconfiguration and Deinitialization before the next RECFB is triggered. By using the QI and QO data variables of the RECFBs, the internal sequences of the RECFBs communicate in a very simple manner. If one sequence fails, the next step will not be executed. Of course this communication can be enhanced according to the reconfiguration purpose. For instance, the coupling of two or more RECFBs is also possible.

The presented engineering methodology handles the two main tasks of reconfiguration applications according to Figure 1 in the following way:

- It performs runtime adaptations on the control application of a distributed automation system, corresponding to the upper half of Figure 1. This includes transfer of state information from one system state to another, if necessary to avoid disruption to the system.
- It performs runtime–checks during execution to ensure correct operation. In case of failures at least the system must be left in a tolerable state, this may include undo functionality for abortive reconfiguration steps. Such failure recovery strategies also require backup and restore mechanisms for state information, corresponding to the lower half of Figure 1. In our approach these topics are handled on different lev-

Figure 10. Regions of interest during the reconfiguration process



els. First of all within one reconfiguration application (an ROI represented by an RECFB) the issues of recovery or undo have to be handled internally of a RECFB. Second also within reconfiguration control applications this issue has to be modeled explicitly. The example in Figure 10 only provides a very simple failure handling. If an error occurs in one of the RECFB instances (i.e. instance names ROI_1, ROI_2 and ROI_3) then the QO data outputs (i.e. RINT_QO, RECONF_QO and RDINIT_QO) are set to false. This is an indication that a failure has occurred and the reconfiguration of a ROI will not take place. But of course more sophisticated error handling procedure can be defined as this simple once.

Another important aspect for reconfiguration is real-time constrained execution. IEC 61499 leaves issues such as timeliness open and delegate's appropriate specifications to compliance profiles and runtime implementations respectively. There exists no barrier to apply such a capability of the runtime environment to the presented engineering methodology. So if the underlying runtime system features timeliness (i.e. deterministic behavior), it can also be applied to the reconfiguration process (doing so certainly the complexity is further increasing).

Reconfiguration Patterns

Based on the above presented reconfiguration method enhanced system reconfiguration the following reconfiguration patterns have been defined in order to provide design rules for reconfiguration processes to the control engineer. These patterns cover the following scenarios:

- Moving of function block instances between IEC 61499 Resources at one device,
- Moving of function blocks instances between different IEC 64199 Devices,
- Moving of resource configurations between IEC 61499 Devices,
- Exchange of Function Block types (same and different interface), and
- Exchange of Resource types.

All of these patterns have in common that they are using RECFBs to describe the reconfiguration process.

Fault Handling During the Reconfiguration Process

In order to manage possible errors or faults during the reconfiguration process the following error scenarios for the zero-downtime system reconfiguration have been covered:

- **Creation of the reconfiguration control application fails:** This situation can be caused by download errors and initialization errors during the deployment phase of the reconfiguration control application. Download errors comprise connection faults, i.e. loss of connection from the engineering environment to the device management (possible reasons might be network or device failure), and device management exceptions.
- **RINIT phase of a reconfiguration step fails:** This situation again comprises connection faults, device management exceptions and initialization errors. Possible reasons are for creating the reconfiguration control application, except network failure, since the reconfiguration application or parts of it are executed locally on the Device.
- **RECONF phase of a reconfiguration step fails:** This category consists of reconfiguration errors and application faults. Reconfiguration errors again comprise

connections faults and device management exceptions. Application faults can be characterized as follows: all reconfiguration tasks terminated successfully, but the overall behavior of the new application is faulty.

- **RDINIT phase of a reconfiguration step fails:** The reasons why the RDINIT phase can fail are quite similar to those of the RINIT phase in. The only differences are that device management exceptions can occur upon deleting FBs or connections (reasons remain the same) and that initialization errors are replaced by de-initialization errors possible for all FBs requiring an INIT-event.

In order to handle these error scenarios the following error handling and fallback strategies have been developed:

- **Error handling during download of the reconfiguration application (RINIT phase):** The error scenarios and therefore the possible handling strategies are the same for these two phases. Connection faults might be recoverable by a simple RETRY; since these errors are time-uncritical it could also prove useful to wait for a short time before retry as a busy network might recover meanwhile. Unrecoverable connection faults terminate the whole reconfiguration process, cleanup is thereby impossible for devices with unreachable manager. Initialization errors can be handled likewise, except that upon failed RETRY of INIT+ the targets should be cleaned up. Device management errors will not be automatically recoverable in general. Already a single unrecoverable error terminates the whole reconfiguration process. Two approaches seem feasible: (i) Terminate upon the first error, which is primitive but rather safe; (ii) Let the user interact (e.g. let him download missing FB types or delete instances already present) and continue if the error could be resolved that way, which is advanced, but dangerous. Upon any termination the targets should be cleaned up.
- **Error handling during RECONF phase:** First of all, error handling during RECONF is time-critical. Upon connection faults two cases must be distinguished: (i) No structural change has been made to the application (only transfer of internal states has occurred so far). If the device is still operational no damage has been caused, otherwise within modern plants emergency measures will be activated if the failed device is a critical one. RETRY, although possible, is a dangerous option to handle connection faults during RECONF, since a faulted connection indicates problems; (ii) At least one structural change has been made already. The system is in an inconsistent state, RETRY is imperative (try to reach a consistent state by all means). If RETRY fails this is the worst case, all that can be done is give alarm and let emergency measures take control.

Also application faults can be divided in two categories: (i) Uncritical faults, e.g. the process are under control in desired manner, but a display is not updated. In this situation it is possible to keep the new application, perform RDINIT as scheduled and resolve the remaining problem by means of another reconfiguration. Certainly error handling strategies below might also be applied here, in doubt the user should be notified to decide how to continue; (ii) Critical faults → the new application is unusable, control over the process might be lost. Several possibilities exist to cope with this situation: UNDO of all changes until a tolerable state is reached; pre-modeled scenarios in the reconfiguration control application could enable reaching a "failsafe" state by further reconfiguration steps or the new application already contains alternative branches along with appropriate switching logic. If the favored branch fails, the second-best solution is tried and so on. Note: this is not real error handling anymore; such a fallback strategy rather belongs to the domain of fault tolerance of the (new) application.

Device management exceptions during RECONF can be treated like critical application faults. In general the RDINIT phase cannot be executed as scheduled if critical errors have occurred during RECONF. Hence the RDINIT phase must either be blocked or adjusted properly (e.g. FB instances planned for removal must not be deleted if still active due to UNDO).

- **Error handling during RDINIT:** Error handling in this phase is quite similar to the RINIT phase, with a few exceptions: errors do not terminate the whole process anymore (e.g. continue with cleanup measures, even if a certain FB could not be deleted, as long as it is safe). De-initializing errors are also candidates for RETRY; caution is advised if de-initializing of an FB fails. In this case it might cause the system to become unstable if such a FB is deleted.

TESTS AND CONCEPT VALIDATION

Engineering Prototype

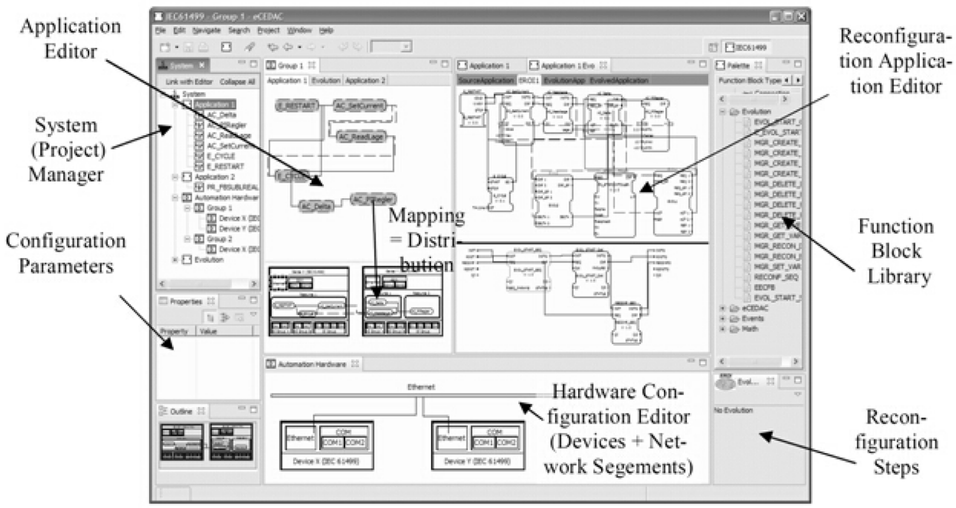
The following section gives an overview of a prototypic implementation of the proposed reconfiguration engineering method. Based on Eclipse SDK and the Graphical Editing Framework (GEF) (Gamma, 2003) the engineering environment supporting the definition of RECFBs is realized as a set of different plug-in for Eclipse. Eclipse is an open source community whose projects are focused on providing an extensible development platform and application frameworks for building software which is very suitable for the engineering of the proposed reconfiguration approach. Figure 11 shows the prototypic implementation.

The main parts of the tool are the Application, the Hardware Configuration Editor, the System (Project) Manager, the Reconfiguration Application Editor and the Function Block and Device Libraries. The Application Editor is used to specify and define function block networks (i.e. IEC61499 applications). Usability and aid for the user are main goals of the editor. Functions like copy/paste or undo/redo are supported by the tool. Checks during drawing of connections are performed in order to avoid programming errors in a very early stage. Functions like zooming helps the user developing larger function block networks. The Outline View provides an overview of the developed control application. Furthermore connections are drawn by a dotted line if connected function blocks are mapped to different devices. This shows the user that communication links have to be configured (see Figure 11).

The Hardware Configuration Editor is responsible for the modeling of the automation hardware (incl. communication network and devices as defined by the IEC 61499 standard). The network can be split into more parts for structuring. Each of these parts can be opened with the editor. Within this editor a function block can be mapped to a device. Mapping means that a function block gets assigned to a device or to a resource if the device has more resources (execution environments for function blocks according to IEC 61499). The system manager gives an overview of different automation projects and can be use to navigate through these projects.

Within the Reconfiguration Application Editor the reconfiguration control applications are modeled based on existing control applications according to the approach presented in this chapter. The tool supports the control engineer in the definition of the reconfiguration control application. It shows the old and the new application and provides also a Delta Analysis as shown in Figure 8. In addition the tool provides also a support to partly generate

Figure 11. Engineering Prototype



a RECFB based on some user interaction. For example if a control engineer wants to exchange a function block of a control application he has to drop the new function block over the old one in the Reconfiguration Application Editor. The tool replaces the function block and generates a RECFB and makes also some configurations of this block based on the reconfiguration templates as explained in the previous section.

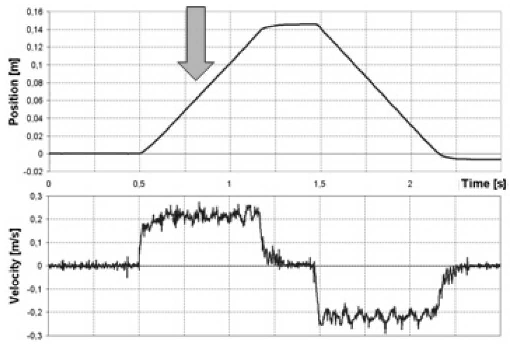
Servo Drive Controller Configuration

First tests in the "Testbed for Highly Distributed Control" for the proposed reconfiguration concept have been performed at the Odo Struger Laboratory of the Automation and Control Institute (ACIN), Vienna University of Technology. These tests gave us the possibility of validating our reconfiguration concepts. In detail, reconfiguration of a prototypic implementation of an execution environment with enhanced (re-)configuration services as explained above, a linear servo drive, has been performed during operation, in this example the movement of the axis according to a position profile.

The control program for the linear servo drive consists of the velocity closed-loop control and the position closed-loop control. The tests have been provided for both of these two parts of the control program. For instance a proportional control algorithm has been exchanged by a new one including also an integral part or in addition a derivative part. The scenario for the velocity closed-loop control was the exchange of the control algorithm without changes in the control hardware. So the ROI for this scenario was restricted to the FB including the control algorithm and its surrounding. During the RECONF sequence an appropriate algorithm has to be applied to avoid disturbances, in this example the output fitting method has been used (Guler, 2003). In the second scenario the position closed-loop control has been exchanged and also the hardware configuration has been changed. An additional controller has been introduced and the new position controller has been applied to the new controller. The results of the second scenario are presented in Figure 12.

The linear axis performs a ramp. During the ramp up sequence the reconfiguration takes place, depicted by the arrow in Figure 12. In the lower part of Figure 12 the corresponding velocity of

Figure 12. Reconfiguration with output-fitting at time 0.75 s



the linear axis gives an impression of the performance of the reconfiguration. There is no influence visible in the velocity graph (of course also in the position graph), although the control algorithm of the position controller has been changed and the position controller itself has been moved to another device within the network. This first result provides first evidence that the proposed reconfiguration approach is sufficient for providing dynamic reconfiguration of lower level control applications in distributed IPMCS. In the provided example the reconfiguration took place only in a few microseconds which is normally fast enough for control tasks in IPMCS.

Communication Reconfiguration

Additional tests for the proposed reconfiguration concept have been also performed by PROFACTOR. To show how a controlled reconfiguration, as explained above, works an example has been created that uses two embedded control boards (DIGI ConnectME development boards, DIGI, 2009) called DIGI1 and DIGI2. Both embedded boards are equipped with the Net+OS 6.3 real-time operating systems (which is based on the ThreadX open source project) and an IEC 61499 compliant runtime environment with enhanced reconfiguration services. The initial control application performs the following steps:

1) A HMI device (e.g. PC) sends a user input (a number) to the DIGI1 device over Ethernet.
2) DIGI1 device sends data to DIGI2 device over Ethernet.
3) The DIGI2 device sends the data to the HMI device over Ethernet and this device displays the result to the user

All communication in this application is over Ethernet, but this can be changed online by a reconfiguration application. This application has been designed such that it changes Step 2 from a communication over Ethernet to a communication over a serial link (for hardware layout see Figure 13).

Figure 14 shows the part of the application that is mapped to the DIGI1 device. The simplified reconfiguration application has to change the PUBLISH_1 FB instance to a SERCOMM_1_0 FB instance and performs the following steps:

Figure 13. Two DIGI ConnectME embedded controller connected via serial link

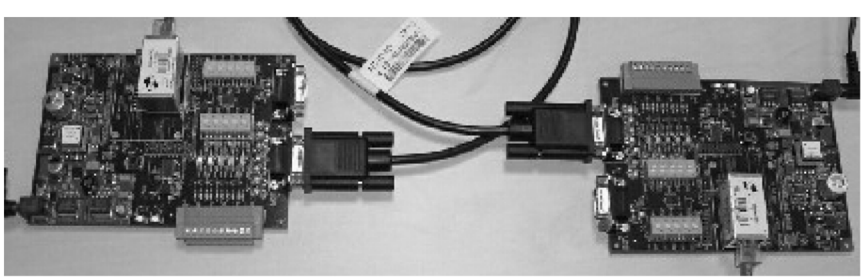
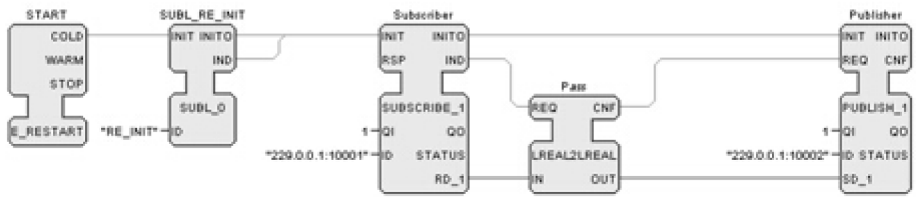
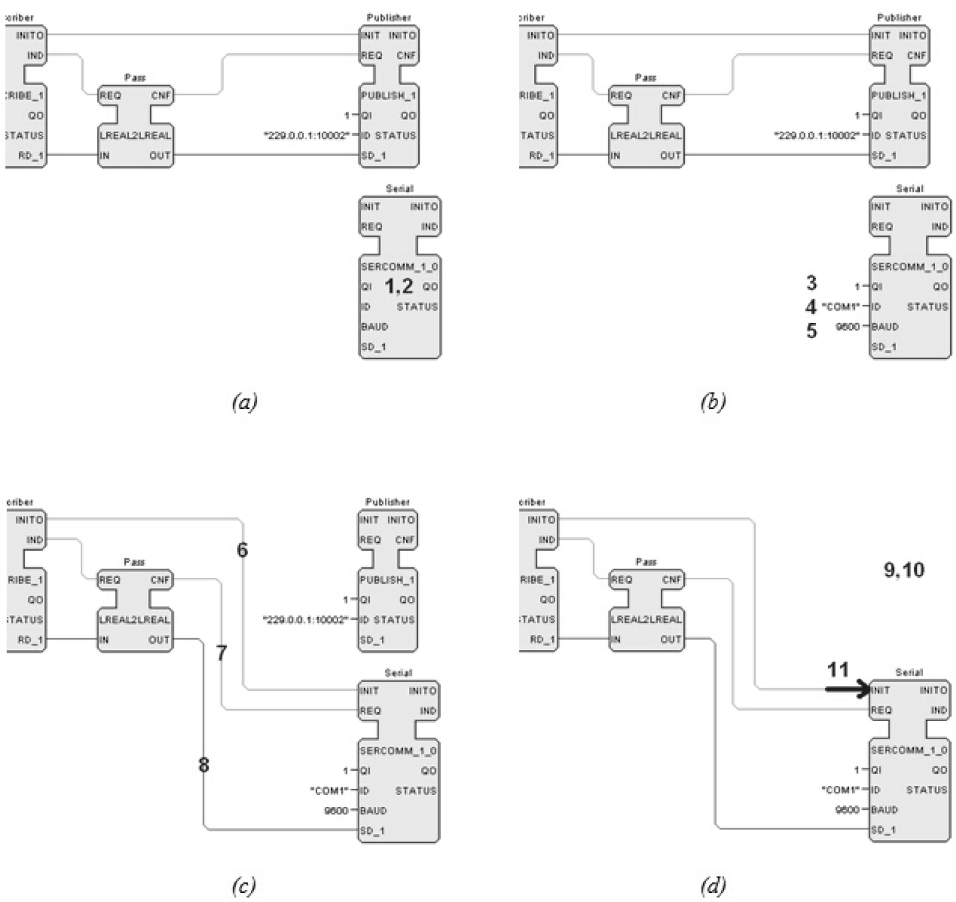


Figure 14. Part of the application in the DIGI1 device



1) Create SERCOMM_1_0 FB with name "Serial"
2) Start "Serial"
3) Set parameter "Serial.QI" to value "1"
4) Set parameter "Serial.ID" to value "COM1"
5) Set parameter "Serial.BAUD" to value "1200"
6) Reconnect[1] "Publisher.INIT" to "Serial. INIT"
7) Reconnect "Publisher.REQ" to "Serial. REQ"
8) Reconnect "Publisher.SD_1" to "Serial. SD_1"
9) Stop "Publisher"

Figure 15. Reconfiguration steps in the DIGI1 device



10) Delete “Publisher”
11) Send INIT event to “Serial”

All these steps can be mapped to the above mentioned special management FBs that are available. Figure 15 shows these steps graphically. Similar steps have to be taken to change the part of the application in device DIGI2.

FUTURE RESEARCH DIRECTIONS

The next steps in our research work are to implement reconfiguration services in an advanced IEC 61499 based runtime environment which is capable of executing reconfiguration control applications. Furthermore a validation of the above described approach will be carried out over a wide range of different control scenarios and applications in the future. Moreover we will also focus our work on an enhanced engineering environment supporting the design and implementation of reconfiguration control applications.

CONCLUSION

This chapter presents an approach for structured modeling of reconfiguration control applications of IPMCS based on the IEC 61499 standard. It has been shown that the concept and elements of the IEC 61499 standard's management model, with little extensions, are sufficient for dynamic reconfiguration of distributed applications.

REFERENCES

Abele, E., Wörn, A., Stroh, C., & Elzenheimer, J. (2005). *Multi Machining Technology Integration in RMS*. Proceedings of the CIRP sponsored 3rd Conference on Reconfigurable Manufacturing, University of Michigan, Ann Arbor, MI, USA.

Brennan, R. W., Fletcher, M., & Norrie, D. H. (2002). An agent-based approach to reconfiguration of real-time distributed control systems. *IEEE Journal on Robotics and Automation*, *18*(4), 444–451. doi:10.1109/TRA.2002.802211

Brennan, R. W., Zhang, X., Xu, Y., & Norrie, D. H. (2002). A reconfigurable concurrent function block model and its implementation in real-time java. *Integrated Computer-Aided Engineering*, *9*, 263–279.

Brinkschulte, U., Schneider, E., & Picioragá, F. (2005). Dynamic real-time reconfiguration in distributed systems: timing issues and solutions. In *Proceedings of the Eighth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing, ISORC 2005*, Seattle, WA, (pp. 174–181).

Burmester, S. (2004). Modeling reconfigurable mechatronic systems with mechatronic UML. *MDAFA - Model Driven Architecture: Foundations and Applications*.

Bussmann, S., & McFarlane, D. (1999). *Rationales for Holonic Manufacturing Control*. In 2nd International Workshop on Intelligent Manufacturing Systems. New York: Leuven.

Christensen, J. H. (2009). *Holobloc Inc. - Resources for the New Generation of Automation and Control*. Retrieved November 2, 2009 from http://www.holobloc.com

DIGI. (2009). *DIGI Connect ME development boards*. Retrieved November 2, 2009 from http://www.digi.com

European Commission. (2004). *MANUFUTURE A Vision for 2020*. Report of the High-Level Group.

Gamma, E., & Beck, K. (2003). *Contributing to Eclipse: Principles, Patterns, and Plugins* (1st ed.). Boston: Addison-Wesley Professional.

Guler, M., Clement, S., Wills, L. M., Heck, B. S., & Vachtsevanos, G. J. (2003). Transition management for reconfigurable hybrid control systems. *IEEE Control Systems Magazine*, *23*(1). doi:10.1109/MCS.2003.1172828

International Electrotechnical Commission - IEC. (2003). *IEC 61131-3: Programmable controllers - Part 3: Programming languages,* (2nd Ed.). Geneva: IEC Press, International Standard. kirchner SOFT GmbH. (2009). *logiCAD - The IEC 61131 Technology Platform*. Retrieved November 2, 2009 from http://www.logicals.com

International Electrotechnical Commission - IEC. (2005). *IEC 61499: Function Blocks - Part 1,2, &3* (1st ed.). Geneva: IEC Press, International Standard.

Koren, Y., Heisel, U., Jovane, F., Moriwaki, T., Pritshow, G., Ulsoy, G., & van Brussel, H. (1999). Reconfigurable Manufacturing Systems. [CIRP]. *Annals of the International Institution for Production Engineering Research*, *48*(2), 527–539.

Lastra, J. Martinez, Lobov, A. Godinho, & L. Nunes, A. (2004). *Function Blocks for Industrial-Process Mesaurement and Control Systems. Institute of Production Engineering*, Report 67. Tampere University of Technology.

Lewis, R. W. (2001). *Modeling control systems using IEC 61499*. Washington, DC: IEE Publishing.

Lorentz, K. (2006). *Torero newsletter no. 2, TORERO Consortium*. Retrieved March 3, 2006 from http://www.uni-magdeburg.de/iaf/cvs/torero

Mens, T., Wermelinger, M., Ducasse, S., Demeyer, S., Hirschfeld, R., & Jazayeri, M. (2005). Challenges in software evolution. In *Proceedings of the Eighth International Workshop on Principles of Software Evolution*, (pp. 13–22).

Rasche, A., & Polze, A. (2005). Dynamic reconfiguration of component-based real-time software. In *Proceedings of the 10th IEEE International Workshop on Object-Oriented Real-Time Dependable Systems, WORDS 2005,* Sedona, Arizona, (pp. 347–354).

Schnakenbourg, C., Faure, J.-M., & Lesage, J.-J. (2002). Towards IEC 61499 function block diagrams verification. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Nashville, TN.

Stanica, M., & Guèguen, H. (2004). Using Timed Automata for the Verification of IEC 61499 Applications. In *Proceedings of the Workshop on Discrete Event Systems*, Reims, France.

Strasser, T., Zoitl, A., Auinger, F., & Sünder, C. (2005). *Towards engineering methods for reconfiguration of distributed automation systems based on the reference model of IEC 61499*. Proceedings of the 2nd International Conference on Industrial Applications of Holonic and Multi-Agent Systems. Copenhagen, Denmark, Aug. 22–24.

Sünder, C., Zoitl, A., Favre-Bulle, B., Strasser, T., Steininger, H., & Thomas, S. (2006). Towards reconfiguration applications as basis for control system evolution in zero-downtime automation systems. *Proceedings of the IPROMS NoE Virtual International Conference on Intelligent Production Machines and Systems, IPROMS*, *2006*(June), 3–14.

Thramboulidis, K., Doukas, G., & Frantzis, A. (2004). Towards an implementation model for FB-based reconfigurable distributed control applications. In *Proceedings of the 7th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing*, Vienna, Austria, (pp. 193–200).

Vyatkin, V., & Hanisch, H.-M. (1999). *A Modeling Approach for Verification of IEC1499 Function Blocks using Net Condition/Event Systems*. Proceedings of IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 1999, Barcelona, Spain.

Wermelinger, M. (1997). A hierarchic architecture model for dynamic reconfiguration. In *Proceedings of the Second International Workshop on Software Engineering for Parallel and Distributed Systems*, Boston, (pp. 243–254).

Zoitl, A. (2008). *Real-Time Execution for IEC 61499*. Research Triangle Park, NC: International Society of Automation - ISA, ISA Press.

ADDITIONAL READING

Al-Safi, Y., & Vyatkin, V. (2009). Ontology-based Reconfiguration Agent for Intelligent Mechatronic Systems in Flexible Manufacturing. *International Journal of Robotics and Computer Integrated Manufacturing*.

Black, G., & Vyatkin, V. (2010). Intelligent Component - based Automation of Baggage Handling Systems with IEC 61499. *IEEE Transactions on Automation Science and Engineering*.

Brennan, R., Vrba, P., Tichy, P., Zoitl, A., Sünder, C., Strasser, T., & Marik, V. (2008). *Developments in Dynamic and Intelligent Reconfiguration of Industrial Automation. Computers in Industry – An International, Application Oriented Research Journal*. Elsevier Editorial.

Cengic, G. (2009), Fuber - FUnction Block Execution Runtime. Retrieved November 2, 2009 from http://sourceforge.net/projects/fuber

Christensen, J. H. (2009), IEC 61499 - A Standard for Software Reuse in Embedded, Distributed Control Systems. Retrieved November 2, 2009 from http://knol.google.com/k/james-christensen/iec-61499#

4DIAC Consortium. (2009), 4DIAC - Framework for Distributed Industrial Automation and Control. Retrieved November 2, 2009 from http://www.fordiac.org

Dubinin, V., & Vyatkin, V. (2009). *Refactoring of Execution Control Charts in Basic Function Blocks of the IEC 61499 Standard*. IEEE Transactions on Industrial Informatics.

Ferrarini, L., & Lorentz, K. (2003). *A case study for modeling and design of distributed automation systems*. Procedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Port Island, Kobe, Japan. pp. 1043–1048.

Fletcher, M., & Norrie, D. (2001). *Realtime Reconfiguration using an IEC 61499 Operating System*. Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS'01) Workshops. IPDPS'01. San Francisco. CA, USA.

Higgins, N., Vyatkin, V., Nair, N., & Schwarz, K. (2009). *Intelligent Decentralised Power Distribution Automation with IEC 61850, IEC 61499 and Holonic Control*. IEEE Transactions on Systems, Machine and Cybernetics, Part C.

Marik, V., Brennan, R., & Pechoucek, M. (Eds.). (2005). *Holonic and Multi-Agent Systems for Manufacturing*. Proceedings of the 2nd International Conference on Industrial Applications of Holonic and Multi-Agent Systems. HoloMAS 2005. Vol. 3593, Lecture Notes in Computer Science. Springer. ISBN: 978-3-540-28237-2.

Marik, V., McFarlane, D., & Valckenares, P. (Eds.). (2003). *Holonic and Multi-Agent Systems for Manufacturing*. Proceedings of the 1st International Conference on Industrial Applications of Holonic and Multi-Agent Systems. HoloMAS 2003. Lecture Notes in Computer Science. Vol. 2744. Springer. ISBN: 978-3-540-40751-5.

Marik, V., Strasser, T., & Zoitl, A. (Eds.). (2009). *Holonic and Multi-Agent Systems for Manufacturing*. Proceedings of the 4th International Conference on Industrial Applications of Holonic and Multi-Agent Systems. HoloMAS 2009. Lecture Notes in Computer Science. Vol. 5696. Springer. ISBN: 978-3-642-03666-8.

Marik, V., Vyatkin, V., & Colombo, V. (Eds.). (2007). *Holonic and Multi-Agent Systems for Manufacturing*. Proceedings of the 3rd International Conference on Industrial Applications of Holonic and Multi-Agent Systems. HoloMAS 2007. Lecture Notes in Computer Science.Vol. 4659. Springer. ISBN: 978-3-540-74478-8.

O³neida. (2009a). IEC 61499 Compliance Profile - Execution Models of IEC 61499 Function Block Applications. Retrieved March 2, 2009 from http://www.oooneida.org/standards_development_Compliance_Profile.html

O³neida. (2009b). O³neida - Network of Networks to Advance Distributed Industrial Automation. Retrieved November 2, 2009 from http://www.oooneida.org

Sardesai, A. R., Mazharullah, O., & Vyatkin, V. (2006). *Reconfiguration of mechatronic systems enabled by IEC 61499 function blocks*. Proceedings of the Australasian Conference on Robotics and Industrial Automation. ACRA 2006. Auckland, New Zealand.

Strasser, T., Fessl, K., Hämmerle, A., & Ankerl, M. (2005). *Rapid Reconfiguration of Machine-Tools for Holonic Manufacturing Systems*. Proceedings of the 16th IFAC World Congress. Prague, Czech Republic.

Strasser. T., Rooker, M., & Ebenhofer, G. (2010). An IEC 61499 Distributed Control Concept for Recongurable Robots. Accepted for publication. *International Journal Computer Aided Engineering and Technology*.

Sünder, C., Vyatkin, V., & Favre-Bulle, B. (2008). *Proving Feasibility and Estimating Parameters of Dynamic Reconfiguration in IEC 61499 Controllers using Formal Modelling. International Journal of Mechatronics and Manufacturing Systems*. IJMMS.

Thramboulidis, K. (2009), CORFU ESS - Engineering Support System. Retrieved November 2, 2009 from http://seg.ece.upatras.gr/Corfu

Vyatkin, V. (2007). *IEC 61499 Function Blocks for Embedded and Distributed Control Systems Design. International Society of Automation - ISA*. ISA Press.

Vyatkin, V. (2009). *The IEC 61499 Standard and its Semantics*. IEEE Industrial Electronics Magazine.

Zoitl, A., Auinger, F., Vyatkin, V., & Martinez Lastra, J. L. (2004). *Towards basic real-time reconfiguration services for next generation zero-downtime automation systems*. Proceedings of the International IMS Forum 2004: Global Challenges in Manufacturing. Como, Italy.

Zoitl, A., Strasser, T., Hall, K. H., Staron, R., Sünder, C., & Favre-Bulle, B. (2007). *The Past, Present, and Future of IEC 61499*. Proceedings of the 3rd International Conference on Industrial Applications of Holonic and Multi-Agent Systems. HoloMAS 2007. Regensburg, Germany.

Zoitl, A., Strasser, T., Sünder, C., & Baier, T. (2009). *IEC 61499 and IEC 61131-3 in Perfect Harmony?* IEEE Industrial Electronics Magazine.

KEY TERMS AND DEFINITIONS

ACIN: Automation and Control Institute.
EC: Embedded Controller.
ECC: Execution Control Chart.
FB: Function Block.
FBN: Function Block Network.
GEF: Graphical Editing Framework.
IPC: Industrial PC.

IEC: International Electrotechnical Commission.

I/O: Input/Output.

IEC 61131: Standard for PLCs.

IEC 61499: Reference model for distributed and reconfigurable automation and control systems.

IPMCS: Industrial Process Measurement and Control Systems.

PC: Personal Computer.

PLC: Programmable Logic Controller.

QoS: Quality of Service.

RCA: Reconfiguration Control Application.

RECFB: Reconfiguration Execution Control Function Blocks.

Reconfiguration: Arrange into a new PLC/EC configuration.

ROI: Region of Interest.

SDK: Software Development Kit

UML: Unified Modeling Language.

XML: Extended Mark-up Language.

ENDNOTE

[1] Reconnect in this context means disconnect the event connection from "Publisher.INIT" and connect it to "Serial.INIT"

This work was previously published in Reconfigurable Embedded Control Systems: Applications for Flexibility and Agility, edited by Mohamed Khalgui and Hans-Michael Hanisch, pp. 55-81, copyright 2011 by Information Science Reference (an imprint of IGI Global).

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

T

U

V

W

Z